Molecular Characterization of *Jatropha curcas* Resources and Identification of Population-Specific Markers

**Anna Ricci · Konstantin Chekhovskiy ·
Perumal Azhaguvel · Emidio Albertini ·
Mario Falcinelli · Malay Saha**

Abstract *Jatropha curcas* L. is cited as one of the best candidates for future oil and biodiesel production. It is widespread in many tropical and subtropical countries but has not yet received much genetic improvement. The objective of this study was to collect *Jatropha* germplasm and characterize it with molecular markers. A total of 64 genotypes, collected from seven geographic locations on two continents, were analyzed with 32 simple sequence repeat and two candidate gene-specific primers (*ISPJ-1* gene and Curcin-P2 gene promoter). In general, markers were found to be highly conserved, and many (40%) were monomorphic in the studied populations. Polymorphic primers, which amplified population-specific fragments, were identified. The polymorphic information content of the polymorphic markers ranged from 0.03 to 0.47. Genetic similarity analysis identified two distinct groups at 0.73 DICE similarity coefficient. Group I included germplasm collected from the islands of Cuba and Cape Verde, and group II consisted of Brazil, Mozambique, and Senegal populations. Island genotypes were found to be very distinct compared to their mainland counterparts. Sequencing of monomorphic fragments identified single nucleotide polymorphism (SNP) between these two groups. High-resolution melting analysis of the SNP in the *Jcps9* locus further confirmed the two gene pools. Sequencing of polymorphic fragments of the *Jc03* locus identified a deletion in a $(GT)_4$ repeat motif in the genotypes in group II. Several population-specific microsatellites and SNP markers have been recognized. The distinct *Jatropha* genotypes and the population-specific molecular markers identified in this study will be valuable resources in breeding programs.

Introduction

The growing demand for biofuels has created an interest in finding new and promising energy species, characterized by high production and environmental sustainability. *Jatropha curcas* L. ($2n=2x=22$) is a promising energy species for its high yield and high-quality oil [14], tolerance to drought, role in soil erosion control, and possible contribution to the economic development of rural and disadvantaged areas [12]. *Jatropha* belongs to the Euphorbiaceae family and is a perennial small tree or scrub plant that produces capsules with usually three seeds [9]. All parts of the seeds could be used for energy production. The oil from kernels (50–55%) is suitable for biodiesel production, and the panel cake, a byproduct of oil extraction, could be burned or used in biogas production [14]. This species not only has direct economic benefits but is also used for making soap and as a barrier for food crops from animal browsing. Some parts of the plant and seeds have curative properties against human

A. Ricci · E. Albertini · M. Falcinelli
Department of Applied Biology, University of Perugia,
Borgo XX Giugno 74,
06121 Perugia, Italy

K. Chekhovskiy · P. Azhaguvel · M. Saha (✉)
Forage Improvement Division,
The Samuel Roberts Noble Foundation,
2510 Sam Noble Parkway,
Ardmore, OK 73401, USA
e-mail: mcsaha@noble.org

diseases [12]. Due to the presence of phorbol esters, *Jatropha* is generally considered toxic; however, some non-toxic genotypes exist in Central America [4, 5, 18, 19].

Jatropha is widespread in the entire tropical and subtropical belts of America, Africa, and Asia. The center of origin is hypothetically defined as Central America [12]. The lack of efforts in genetic improvement of this species has limited its success in many areas where it is adapted [36]. *Jatropha* is still considered a wild species [14], but it could be a vital crop for energy purposes due to its many potential advantages. Varieties with desirable traits such as improved yield and oil quality can be achieved through breeding. Understanding of the extent of genetic variability in *Jatropha* is critical for the success of breeding programs.

Characterizing genetic diversity in a species using molecular markers can generate useful information for breeding programs. *Jatropha* is considered predominantly an out-crossing species. Plants are monoecious with male and female flowers in the same inflorescence [9]. Even though apomixis was reported [7], gametophytic apomixis seems absent in this species [1]. Vegetative propagation is frequently used for crop multiplication [12]. To evaluate genetic variability in such a cross-pollinated species, separate DNA profiles from randomly selected individuals from each population are needed for the assessment of within- and among-population variations [40].

Several studies on genetic variability in *Jatropha* populations were reported using different molecular markers, e.g., amplified fragment length polymorphism, randomly amplified polymorphic DNA (RAPD), single primer amplification reaction, inter-simple sequence repeat (ISSR), and simple sequence repeat (SSR) [1, 4, 5, 15, 23, 27, 37]. DNA marker (RAPD, ISSR, and SSR) analyses revealed that the *Jatropha* seeds commercialized worldwide consisted of few genetically identical or closely related clones and showed very low genetic variation either within or between varieties [1]. A diversity analysis of 192 accessions collected throughout Brazil reflected a common ancestry, genetic drift, and intensive selection of cultivated material since its introduction [29]. Narrow genetic bases were reported among germplasms collected in India [4] and in China [37]. Wider variability was present among genotypes collected from different geographic areas [5] and between toxic and non-toxic genotypes collected in Mexico [23]. Recently, Ovando-Medina et al. [22] summarized the genetic diversity research on *Jatropha* as published in indexed journals. They indicated that breeding for productivity has been restricted by the lack of genetic information and suggested continued research on genetic diversity in domesticated populations.

The single nucleotide polymorphism (SNP) marker is recognized as a powerful marker system, and its polymorphism can be analyzed rapidly with high-resolution melting (HRM) analysis [17]. Taking advantage of the thermodynamic variations of the amplicons, genetic variants with differences in base composition can be detected by their characteristic melting curves [21]. HRM has been used for genotyping, varietal identification, and molecular mapping [8, 17, 35]. In this study, we tested the applicability of a SNP marker in differentiating *Jatropha* germplasm.

The objectives of this study were to assess the genetic diversity and phylogenetic relationships among seven *Jatropha* populations collected from different phytogeographic regions using molecular markers and to identify population-specific markers. Here, we report for the first time the development of a population-specific SNP marker in *Jatropha*.

Materials and Methods

Plant Material

Seeds from seven *Jatropha* populations were collected from different geographical locations of Brazil, Cape Verde, Cuba, Mozambique, and Senegal (three populations). Genomic DNA was extracted from young leaves of eight plants of each population except those of Mozambique from which 16 samples were collected (total of 64 plant samples). All the leaves were ground in liquid nitrogen. For DNA extraction, Genelute Plant Genomic DNA Kit (Sigma-Aldrich Corp., Woodlands, TX, USA) was used for the Brazil, Cape Verde, Cuba, and Senegal populations, and Qiagen DNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA) was used for the Mozambique population. DNA was extracted at the University of Perugia, Italy, from all populations except Mozambique which was collected later and the DNA was extracted at the Noble Foundation, USA. The quality and quantity of DNA were checked on agarose gel and with NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA).

Primers and Touchdown PCR

A total of 42 *Jatropha*-specific SSR primer pairs (PPs) were selected from public resources (Table 1 in "Electronic supplementary material"). Seven additional PPs were designed with Primer3 (http://frodo.wi.mit.edu/primer3/) from *Jatropha* sequences available in the NCBI nucleic acid sequence database (accessions from GQ495139-GQ495168; http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide) (Table 1 in "Electronic supplementary material"). The major parameters for primer design include primer lengths of 18–27 with an optimum length of 20 nucleotides, an optimum annealing temperature of 60°C, GC content between 20% and 80%, and a product size of

100–200 bp. The oligonucleotides were synthesized by Sigma-Aldrich (Sigma-Aldrich Corp., Woodlands, TX, USA) and Operon Technologies (Operon Technologies Inc, Alameda, CA, USA). An 18-bp M13 tail was attached to each forward primer. A Curcin-P2 PP, specific for the promoter region of the curcin gene, and a SCAR PP, ISPJ1, presumed to be associated with the toxicity trait [4, 5], were also included in this study. A BLAST analysis of sequences, from which PPs were designed, was performed (http://www.ncbi.nlm.nih.gov/projects/mapview/static/MVPlantBlast.shtml?10, accessed on Nov. 15, 2009) to identify similarity with the sequences of model plant species.

PCR amplification was performed using GoTaq DNA Polymerase and 5X Colorless GoTaq Reaction Buffer (Promega Corporation, Madison, WI, USA). The 10-µl PCR mix consisted of 50 mM MgCl$_2$, 1.5 mM dNTPs, 0.25 µM forward primer, 1 µM reverse primer, 1 µM fluorescent dye with M13 tail, 0.45 U Promega Taq, and 10 ng genomic DNA. A single touchdown PCR procedure was set up for all primer amplifications [3, 39], and the PCR reactions were performed in the Gene Amp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). The PCR procedure consisted of 3 min at 95°C, followed by six cycles of 45 s at 94°C, 5 min at 68°C (with a 2°C decrease in each cycle), and 1 min at 72°C, followed by eight cycles of 45 s at 94°C, 2 min at 58°C (with a 1°C decrease in each cycle), and 1 min at 72°C, followed by 25 cycles of 45 s at 94°C, 2 min at 50°C, and 1 min at 72°C, followed by a final 7-min extension step at 72°C and stored at 4°C. Duplicate PCR reactions were made for PPs JCSSR-24, Jc07, Jcds41, and Curcin-P2 as a confirmation of the previous data.

Genotyping and Sequencing

The PCR products were analyzed with ABI 3730 DNA Analyzer (Applied Biosystems) following the protocol suggested by Saha et al. [31]. ABI data were scored as present (1) or absent (0) with GeneMapper Software v 3.7. Monomorphic fragments from ten PPs obtained in 16 genotypes (four from Mozambique and two from each of the other six populations) and two polymorphic fragments from the Jc03 locus in 14 genotypes (two from each population) were excised from agarose gel, and DNA was extracted using the DNA Gel Extraction Kit (Millipore-Genomics, Bedford, MA, USA). Purified DNA was used as a template for the second PCR amplification which was performed in 30-µl volume following the same protocol described above. The DNA samples were then sequenced in an ABI 3730 DNA Analyzer (Applied Biosystems) following the manufacturer's protocols.

High-Resolution Melting Analysis

High-resolution melting (HRM) analysis was performed to validate a SNP detected in the *Jcps9* locus, which can be used as a diagnostic SNP marker. LightScanner® System (Idaho Technology Inc., Salt Lake City, UT, USA) was used to analyze ten samples from six populations with two replicates. The annealing temperature of *Jcps9* primer was adjusted based on the results from a gradient PCR with a temperature range from 50°C to 62°C. The 10-µl PCR reaction mix consisted of 1X LightScanner Master Mix (Idaho Technologies), 2.5 µM of forward and reverse primers, and 10 ng of genomic DNA. Mineral oil (25 µl) was added to each well. The PCR procedure included a 2-min initial denature step at 95°C, followed by 40 cycles of 30 s at 94°C and 30 s at 60°C, and a final cycle of 30 s at 94°C and 30 s at 26°C and a final hold at 4°C. The amplification was performed in the Gene Amp PCR System 9700 (Applied Biosystems); the PCR product was then transferred to Lightscanner 384-well system, and a melting cycle was performed by increasing the temperature at 0.1°C s^{-1} from 70°C to 95°C. The melting data were analyzed with the LightScanner Software with CALL-IT 2.0 (Idaho Technologies) using the small amplicon module. The samples were assigned to different groups by the LightScanner Software's hierarchical clustering based on the shape of their normalized melting curves and visualized as normalized melting curve plots, normalized melting peaks (normalized negative first derivative plot), and normalized difference plot. One sample of the BRAC6 population (Fig. 5b) was used as standard in the normalized difference plot.

Data Analysis

Polymorphic information content (PIC) was used to define the amount of information from a marker locus and it was calculated using the equation:

$$\text{PIC}_i = 1 - \sum_{j=1}^{n} P_{ij}^2$$

where PIC$_i$ is the polymorphic information content of marker i, and P_{ij} is the frequency of the jth pattern for the marker i and the sum was made over n patterns [2, 6]. PIC calculations were made on each population and on the combined data across populations (Table 1). Five genotypes (CVA11, CVE15, S2-2, S3-85, M1-2) with many missing values were excluded from analysis.

Analysis of molecular variance (AMOVA) was done using GenAlEx 6.4 (http://online.anu.edu.au/BoZo/GenAlEx/index.php [25]) to partition the genetic variance within populations and among populations. Cluster analysis was performed using the data obtained from 34 primer pairs across 59

Table 1 Polymorphic information content of *Jatropha*-specific primers observed in populations of distinct geographic origins

Primer name	CB+CV[a] populations	Other[b] populations	Total PIC[c]
Curcin-P2	0.39	0	0.32
ISPJ1	0.12	0.16	0.16
Jc04	0.21	0	0.06
Jc07	0.28	0	0.26
Jc08	0	0.12	0.09
Jc10	0.21	0.18	0.19
Jc13	0.39	0	0.35
Jc14	0.33	0.04	0.37
Jcd41	0	0.04	0.03
Jcds10	0.41	0.48	0.47
Jcds24	0.35	0.15	0.23
Jcds66	0.28	0.23	0.24
Jcms 21	0.33	0.08	0.39
Jcps17	0.12	0.18	0.17
Jcps20	0	0.04	0.03
Jcps21	0.21	0.23	0.23
Jcps6	0.12	0.24	0.22
Jcps7	0	0.04	0.03
Jcps9	0	0.20	0.17
JcSSR18	0	0.04	0.32
Jc03	0	0	0.30
Jc05	0	0	0.30
Jc09	0	0	0.30

[a] Population collected from islands of Cuba and Cape Verde

[b] Populations collected from mainlands of Brazil, Mozambique and Senegal

[c] PIC was calculated from 59 genotypes excluding those with many missing values

genotypes (excluding CVA11, CVE15, S2-2, S3-85, M1-2). The similarity matrices and dendrogram were constructed using the Numerical Taxonomy and Multivariate Analysis System (NTSYS-pc) software package, version 2.1 [28]. The genetic similarity among the genotypes was calculated using the DICE similarity coefficient [10]. For clustering, the similarity data were analyzed in the SAHN module using the UPGMA method. The TREE module was used to create the dendrogram. Using the qualitative data of 0 and 1, Jaccard's similarity coefficient was calculated and the data were transformed into a square matrix. The principal coordinate analysis (PCA) by Eigen module was performed using NTSYS-pc, version 2.1 (Applied Biosystematics, Setauket, NY, USA) and visualized similarities among the *Jatropha* accessions. The sequence data were analyzed with DNASTAR software using SeqMan as the alignment tool (Lasergene Inc., Madison, WI, USA). Three PPs failed to produce high-quality sequences and were deleted from the analysis.

Results

Primer Selection and Amplification

Eleven PPs amplified similar products which were considered as duplicates with different primer names assigned by two groups [5, 37]; therefore, only data from one set were included in the final data analysis. Additionally, four primers (Jcmns209, Jcps3, Jc06, and Jc16) with no or very poor amplifications were excluded from the statistical analysis. The final data set includes amplification results from 34 unique PPs. Several PPs were randomly selected and re-amplified in the genotype pool to verify the amplification products. Results from the re-runs showed very similar amplification patterns and thus confirmed reliability of the data. BLAST analysis of *Jatropha* sequences, from which PPs were developed in this study, did not show similarities with known sequences of the model species (data not shown).

The 34 PPs generated 62 fragments with an average of 1.82 fragments per PP. Eleven PPs amplified single products, and they are mostly monomorphic. In general, all the 23 polymorphic PPs in this study can be divided into three distinct groups: (1) PPs polymorphic in both CC and BMS groups (eight primers), (2) polymorphic in either CC or BMS but monomorphic in the other (8 primers), and (3) specific fragments for either CC or BMS groups (seven primers).

The number of fragments amplified in each plant varied from 36 to 44. The average fragment number per individual in each population is 38 in Cape Verde, 40 in Brazil, Cuba, and Senegal, and 41 in Mozambique. These results indicate that the tested PPs were equally effective in all the studied populations. Of the 62 fragments, 25 (40%) were monomorphic and 15 (24%) were population specific, i.e., present or absent in certain populations. The PPs Jc03, Jc05, and Jc09 amplified two fragments each; one was specific for Cuba and Cape Verde (CC group) and the other one for Brazil, Mozambique, and Senegal populations (BMS group). PP JcSSR18 amplified three fragments, two of which were mainly seen in the BMS group and the other was specific for the CC group. PP Jc13 amplified four fragments, one mainly in the CC group with one exception, one specific to BMS groups, and others present in a few genotypes in the CC group. Jc07 and Jcms21 also produced fragments which were mainly found in CC or BMS groups. Curcin-P2 amplified two specific fragments for CC accessions with only one exception.

Polymorphic Information Content

The total polymorphic information content values ranged from 0.03 to 0.47 in 23 polymorphic PPs (Table 1). The

highest PIC value (0.47) was from PP Jcds10. PPs Jc03, Jc05, and Jc09 had a PIC content of 0 in either CC or BMS groups but had a total PIC of 0.30 (Table 1). These PPs amplified fragments specific to either CC or BMS groups. There were several primers which were monomorphic in one group but fairly polymorphic in another group (Curcin-P2, Jc04, Jc07, Jc08, Jc13, and Jcps09) and thus had a PIC of 0 or close to 0 in one group but was fairly informative in the other group. Higher PIC values across populations were observed in Jcds10, Jcds66, Jcds24, Jc10, Jcps21, Jcps17, Jcps6, and ISPJ1, and these PPs are the informative primers for these populations.

Genetic Diversity

The AMOVA of the genetic relations within and among populations revealed that the variations among populations were more pronounced than the variations within populations. The variation within populations was 33% of the total variation, while among population variance contributed 67%. Two distinct clusters were identified (Fig. 1). Cluster I included populations from islands, e.g., Cuba and Cape Verde (CC group), and cluster II included populations from mainland, i.e., Brazil, Senegal, (three populations) and Mozambique (BMS group). The separation of these two
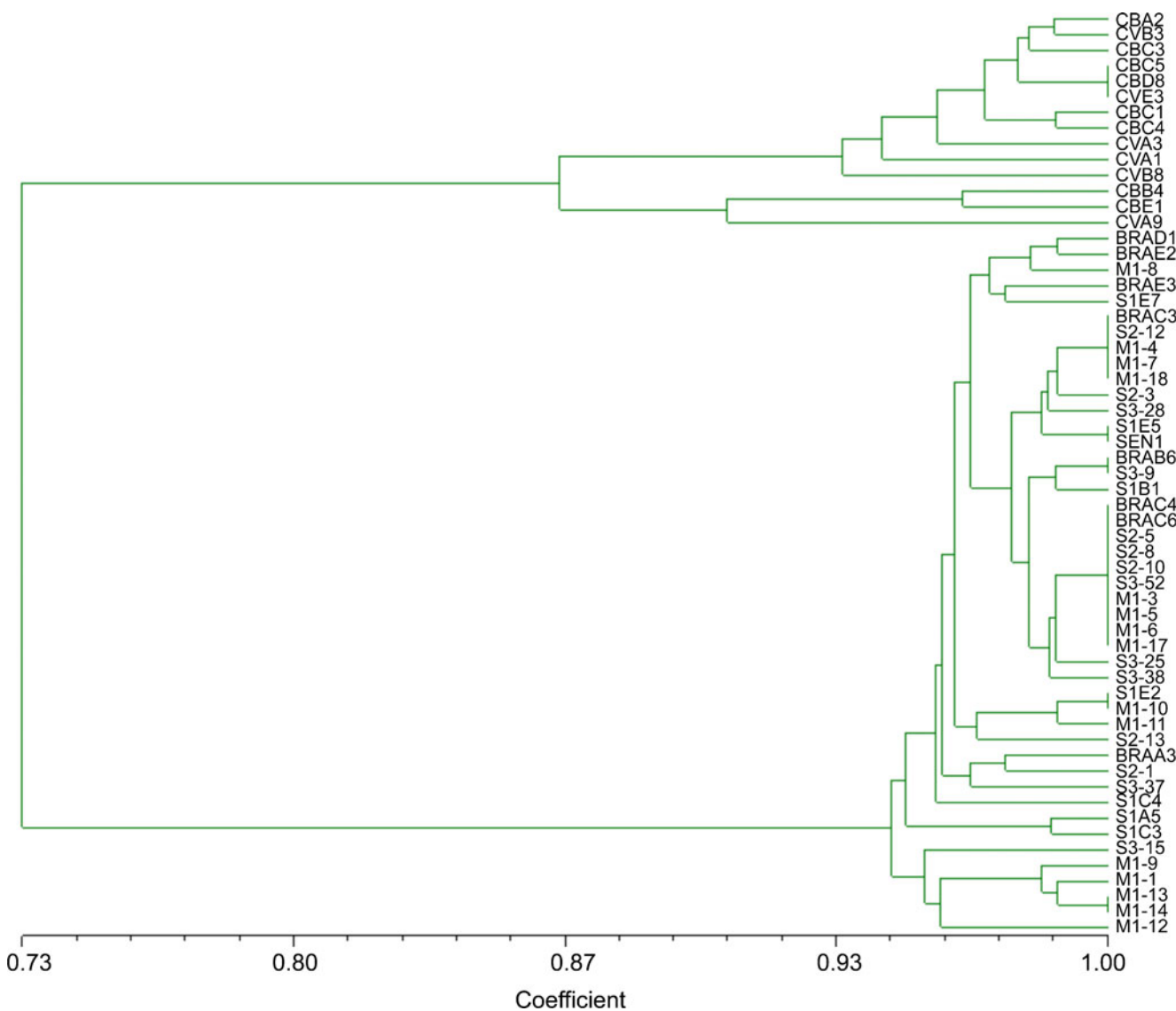

Fig. 1 Dendrogram showing genetic similarity among and within *Jatropha* populations. The origin of an individual is indicated with the first one or two letters in its name where *CB* = Cuba, *CV* = Cape Verde, *BR* = Brazil, *M* = Mozambique, and *S* = Senegal. Microsatellite markers differentiate samples into two distinct clusters. Cluster I consists of genotypes from Cuba and Cape Verde, and cluster II includes genotypes from Brazil, Mozambique and Senegal
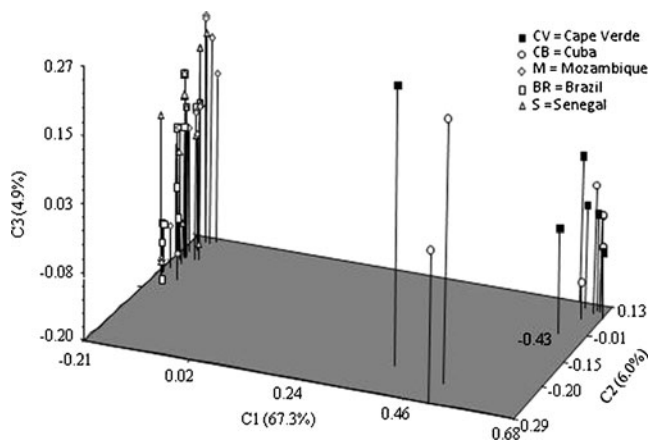
Fig. 2 Three-dimensional plot from PCA of 59 *Jatropha* genotypes analyzed with 62 microsatellite markers. The origin of an individual accession is indicated with *symbols* and details are given in the *legend*

clusters occurred at a similarity coefficient value of 0.73. Within cluster I, two subclusters were separated at 0.87, and both included genotypes collected from Cuba and Cape Verde. Genotypes within the BMS group were very similar with minimum similarity coefficients of 0.94, and genotypes of all five populations were distributed in a homogeneous way. The two major clusters perfectly reflected the populations' broad geographical origins; however, subclusters within either major cluster did not correspond to each population's specific geographical origin.

The PCA among the 59 *Jatropha* genotypes calculated using Jaccard's similarity coefficient agreed with the clustering analysis by forming two distinct groups, BMS and CC (Fig. 2). The first, second, and third components of the PCA explained 67.3%, 6.0%, and 4.9% of the total variance, respectively. The grouping of populations was highly concordant with the dendrogram (Fig. 1), suggesting that molecular data obtained from the current study were robust and reliable.

Sequence Variability

The alignment of 32 forward and reverse fragments from 16 genotypes (two from each population except Mozambique in which four genotypes were sequenced) at six monomorphic loci produced identical sequences. Sequences from the monomorphic locus *Jcps9* were also highly conserved among all genotypes except a SNP at 96 bp (Fig. 3). The cytosine nucleotide (C) in all genotypes of Brazil, Mozambique, and Senegal populations substituted to a thymine nucleotide (T) in the Cuba and Cape Verde genotypes (Fig. 3). Alignment of the 14 sequences from the locus *Jc03* showed an addition of four GT repeat motifs in the microsatellite regions of the genotypes from Cuba and Cape Verde (Fig. 4) which resulted in fragment size polymorphisms. All the genotypes of BMS populations have $(GT)_{11}$ motifs in the microsatellite regions while those of CC populations have $(GT)_{15}$ motifs (Fig. 4).

HRM Genotyping

We used HRM to genotype the SNP found in Jcps9 amplicons. The result from a gradient PCR indicated that 60°C is Jcps9's optimum annealing temperature for HRM analysis. Jcps9 amplicons from ten genotypes from six populations were used in this analysis. Grouping of samples were visualized in two types of plot (Fig. 5a, b). Melting peaks, which are normalized negative first-derivative plots, are shown in Fig. 5a. Figure 5b shows the normalized difference curves, which was recalculated using a sample from BRAC6 as the baseline. In both plots, the samples were separated into two distinct groups where the gray ones belong to the CC group and the red ones belong to the BMS group. The melting temperature (indicated by the peak position in Fig. 5a) of samples from the BMS group is higher than that from the CC group due to the SNP composition. The results were consistent between the two

Fig. 3 Alignment of DNA sequences obtained from 16 *Jatropha* genotypes at the *Jcps9* locus. A single mutation (C/T) at 96 bp, *boldface font*, differentiates Cuba and Cape Verde genotypes from the other populations

```
BRAB6   TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
BRAC3   TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
CBA2    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
CBB4    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
CVA1    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
CVA9    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
S1A5    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
S1B1    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
S2-1    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
S2-3    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
S3-9    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
S3-15   TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
M1-1    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
M1-3    TATTCATTTCTGTGTGTGTGTGTGTGTGTGTGT--------CAATTTGGTTTATATGAGGGTGGGTTGAACTCACAACGTT
```

Fig. 4 Alignment of the microsatellite regions of 14 genotypes at the *Jc03* locus showing the insertion of (GT)$_4$ in Cuba and Cape Verde genotypes

replicates. New genotypes of both CC and BMS groups included in this analysis (which were not sequenced before) separated into their expected groups (Fig. 5b).

Discussion

Jatropha is a multipurpose species with many desirable attributes. The importance of *Jatropha* has increased significantly in recent years because of its use as a potential source of biodiesel. The major constraint for improvement in *Jatropha* is the narrow genetic base [5]. Understanding genetic variability in a species is a prerequisite for initiating any rational breeding program. *Jatropha* is a hardy perennial shrub which is believed to be a native of Mexico and Central America and is widespread throughout the tropical regions of Africa and Asia [12]. Plant populations on islands and mainlands may display local morphological divergence in a short evolutionary time frame [33]. *Jatropha* plants collected from the mainlands of Brazil, Mozambique, and Senegal and from the islands of Cuba and Cape Verde were included in this study. Germplasm collections from distinct geographic groups increased the possibility of finding unique gene pools which can potentially be used in future breeding programs.

Microsatellite or simple sequence repeat markers have proven useful for genetic diversity analysis in different crop species [20, 30]. Only a few SSR markers have been developed for the genetic analysis of *Jatropha* [5, 37]. We used 49 PPs for the genetic variability analysis of *Jatropha* collections. A set of 11 PPs obtained from Basha et al. [5] showed very similar amplifications with another set of primers obtained from Sun et al. [37]. It was observed that though the PPs have different names and/or have different forward and/or reverse primer sequences, they were originated from the identical nucleotide sequences (same gene bank ID; Table 1 in "Electronic supplementary material"); thus, the duplicate set was deleted from final

analysis. All the PPs (49) used in this study were originated from the *Jatropha*-specific sequences, but four of these (8%) did not amplify in the studied genotypes. A very high
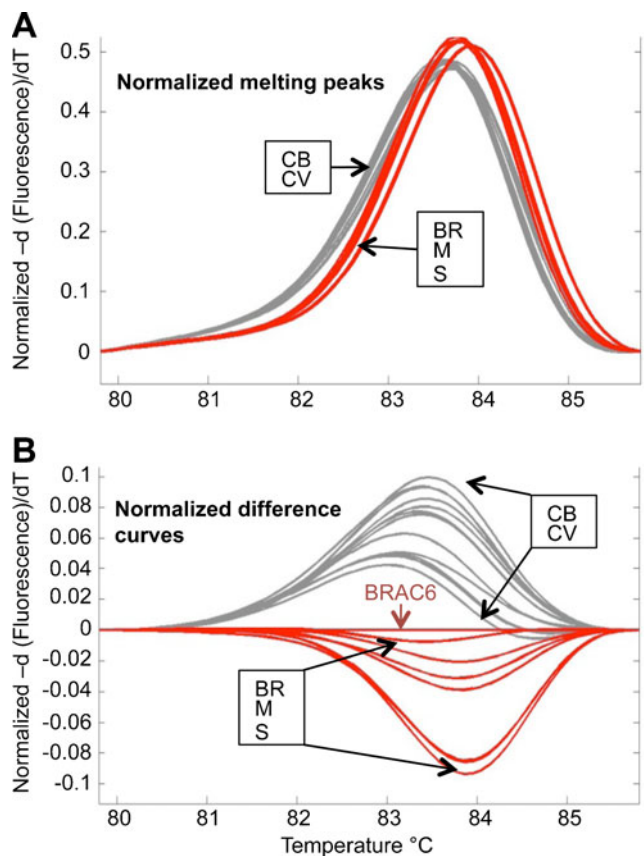

Fig. 5 High-resolution melting analysis of BRAC6, S1C4, S2-8, M1-1, CBA2, CBB4, CBC1, CBC3, CBD8, and CVB8 genotypes (the origin of an individual is indicated with the first one or two letters in its name where *CB* = Cuba, *CV* = Cape Verde, *BR* = Brazil, *M* = Mozambique, and *S* = Senegal) using primers designed from *Jcps9* locus. **a** Normalized melting peaks, **b** normalized difference plot using BRAC6 as baseline. *Red-colored peak* represents genotypes with "C" in the SNP locus (Brazil, Mozambique, and Senegal populations) while the *gray* color represents genotypes with "T" nucleotide which is present in the Cuba and Cape Verde populations

success rate (92%) of the primers across populations indicated high levels of genome conservation in *Jatropha* germplasms. Similar results were also reported by previous investigations [5, 26, 27].

Despite a high level of genomic similarity, the molecular marker profiling showed distinct amplification patterns in the germplasm collections. There were 60% polymorphic fragments, 24% of which are population specific, e.g., either present or absent in certain populations. Thus, in addition to the combined analysis, PIC calculations were also performed in individual groups. Even though three primers (JC03, JC05, and JC09) have PIC values of 0.30 in all populations, their PIC values within CC and BMS groups were 0. There are four other primers which had very similar profiling with one or two exceptions. Basha et al. [5] also reported population-specific markers in their *Jatropha* collections.

The dendrogram based on UPGMA and the PCA analyses revealed two distinct clusters in the germplasm collections. One cluster was comprised of accessions from Cuba and Cape Verde while the other cluster encompassed accessions from Brazil, Mozambique, and Senegal regions. The first components of the PCA itself explained most of the variation (67.3%) in the collections, thus indicating that the two groups are very distinct (Fig. 2). Within the BMS cluster, plants were very similar (coefficient >0.94), and several plants from Brazil, Mozambique, and Senegal regions are genetically identical at the studied loci. Plants within the CC group were comparatively more variable than the BMS group; however, several plants from Cuba and Cape Verde were genetically identical (Fig. 1). Data from AMOVA also suggested similar trends in among- and within-population variations. The genetic closeness of accessions within clusters I and II indicate a common source of origin for these materials.

Genotypes from Cuba and Cape Verde (island populations) were very different from the mainland (BMS) populations. In general, islands would have a biota which would be a subset of that on the continent or would have changed as a result of independent evolution and extinction. Serebryanaya and Shipunov [33] studied the morphological variation in island and mainland plant populations of the north European White Sea and found ten cases where differences were best explained by the existence of short-scale evolutionary processes. It is possible that the *Jatropha* plants in the islands of Cuba and Cape Verde went through a similar evolutionary process and became a distinct gene pool compared to the mainland counterparts. Accessions from Cape Verde had been transported by Portuguese seafarers to different regions in Asia and Africa. This may be the reason for clustering the Cape Verde accession with the Indian germplasm [5]. Clustering together of accessions from different regions indicates lack of gene flow between adjacent populations in each region.

All the studied populations showed least amount of within-population variations. Studies conducted in India, China, and Brazil also reported similar trends in *Jatropha* populations [22]. However, wide within-population variations were observed in Malaysian collections analyzed with morphological data [34]. In spite of being an outcrossing species, *Jatropha* has very limited genetic variability. The low variability can be due to one or all of the following factors: the vegetative propagation, the natural and/or the anthrophic selection, and the management system of local farmers in the cultivated areas. The ease of vegetative propagation can be a limiting factor for the variability in the populations. *Jatropha* is widespread in many areas where it is used as a living fence [12]. At the same time, vegetative propagation may also be considerably common in the cultivated areas where the seeds are the principal product. Seeds of *Jatropha* are mainly used for oil and soap production. We hypothesize that farmers prefer to harvest seed only to produce oil, not to conserve them to reproduce new plants. The populations in many areas could be constituted of only one or a few genotypes (clones) selected by the farmers. If that hypothesis is correct, the variability could be low (or could be decreased) by the actions of humans. In cassava, which is allogamous, but primarily vegetatively propagated, the local farmers of some areas usually reproduce new plants using the vegetative suckers and also include volunteer seedlings in their fields, promoting genetic diversity and landraces [11, 32]. Seeds of wild *Jatropha* are high in phorbol esters and are very toxic to birds, mice, and rats [24], which may also restrict natural seed dispersal. All of these factors together might restrict the genetic variability in this species.

The use of SSR markers in evolutionary and genetic studies does have limitations due to homoplasy when fragments of the same size may not have the same nucleotide sequences which can underestimate the genetic variability that exists in plant populations [38]. Polymorphisms in the forms of insertions, deletions, and substitutions are fairly common within SSR motifs and their flanking regions [30]. Sequences from six loci had revealed no variations, suggesting the solidity of these primers in the microsatellites and in the flanking regions. However, the primer Jcps9 showed a SNP at 96 bp which was consistent between genotypes of CC and BMS groups. Primer Jc03 amplified two alleles 8 bp apart. The longer fragment of 291 bp was specific to CC individuals, and the shorter fragment was specific to BMS groups. Sequencing of these fragments confirmed the deletion of four GT repeats in BMS individuals which was the cause of the polymorphism (Fig. 4). Polymorphic markers to ascertain a population of unknown origin have long been an area of interest to plant breeders. Most markers have alleles which are fairly conserved between each population. Jc03 and Jcps9 PPs can be

potentially used as diagnostic markers to distinguish the mainland populations from the islands studied in this project.

The narrow sampling in this study may limit its applicability across a larger population since only two genotypes were sampled from each population. The SNP variation identified in the *Jcps9* locus was tested in a large set of genotypes using HRM analysis. SNP markers are valuable tools for the genetic analysis of complex traits of agronomic importance and marker-assisted selection. Melting curve analysis using fluorescent dyes has proven to be a highly sensitive method for mutation discovery and SNP genotyping [17]. SNPs are more amenable to high-throughput automated genotyping assays and thus allowed samples to be genotyped faster and more economically than with SSRs [13]. SNP detection technologies have evolved from expensive, time-consuming, and labor-intensive processes to highly automated, efficient, and relatively inexpensive methods of DNA marker detection [16]. HRM genotyping separated the samples with unknown sequences into each expected group and validated the SNP by confirming the samples for which sequences were known. SNPs are highly conserved throughout evolution and within a population; thus, it can be effectively used for genetic linkage analysis, population studies, and studying marker trait associations. All the tested genotypes in the CC and BMS groups showed distinct HRM profiles; thus, *Jcps9* should be considered as a diagnostic marker for these populations.

In conclusion, *Jatropha* germplasms from the islands of the Pacific Ocean and mainlands of Brazil, Mozambique, and Senegal showed a high level of genetic similarity within populations. However, island genotypes were found to be very distinct from their mainland counterparts at several loci. Several microsatellite markers (JC03, JC05, and JC09) and a single base substitution at *Jcps9* locus clearly separated the island and mainland populations. These SSRs and SNP are useful diagnostic markers for distinguishing the island and mainland *Jatropha* populations.

References

1. Ambrosi DG, Galla G, Purelli M, Barbi T, Fabbri A, Lucretti S et al (2010) DNA markers and FCSS analyses shed light on the genetic diversity and reproductive strategy of *Jatropha curcas* L. Diversity 2:810–836
2. Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimising parental selection for genetic linkage maps. Genome 36:181–186
3. Azhaguvel P, Li W, Rudd JC, Gill BS, Michels GJ Jr, Weng Y (2009) Aphid feeding response and microsatellite-based genetic diversity among diploid *Brachypodium distachyon* (L.) Beauv accessions. Plant Genet Resour 7:72–79
4. Basha SD, Sujatha M (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
5. Basha SD, Francis G, Makkar HPS, Becker K, Sujatha M (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
6. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331
7. Chang-wei L, Kun L, You C, Yong-yu S (2007) Floral display and breeding system of *Jatropha curcas* L. Forestry Studies China 9 (2):114–119
8. De Koeyer D, Douglass K, Murphy A, Whitney S, Nolan L, Song Y et al (2010) Application of high-resolution DNA melting for genotyping and variant scanning of diploid and autotetraploid potato. Mol Breed 25:67–90
9. Dehgan B, Webster GL (1979) Morphology and infrageneric relationships of the genus *Jatropha* (Euphorbiaceae). Univ Calif Publ Botany 74:1–72
10. Dice LR (1945) Measures of the amount of ecological association between species. Ecology 26:297–302
11. Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A et al (2003) Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theor Appl Genet 107:1083–1093
12. Heller J (1996) Physic nut, *Jatropha curcas*. Promoting the conservation and use of underutilized and neglected crops 1. IPGRI, Rome, Italy, 1–66. Available online at http://www.bioversityinternational.org/publications/publications/publication/issue/physic_nut_brijatropha_curcasi_l.html
13. Hurley JD, Engle LJ, Davis JT, Welsh AM, Landers JE (2005) A simple, bead-based approach for multi-SNP molecular haplotyping. Nucl Acids Res 32:186
14. Jongschaap REE, Corré WJ, Bindraban PS, Brandenburg WA (2007) Claims and facts on *Jatropha curcas* L. Global *J. curcas* Evaluation, Breeding and Propagation Programme. Report 158. Available online at http://library.wur.nl/way/bestanden/clc/1858843.pdf
15. Kumar RS, Parthiban KT, Rao MG (2009) Molecular characterization of *Jatropha* genetic resources through inter-simple sequence repeat (ISSR) markers. Mol Biol Rep 36:7
16. Kwok PY, Chen X (2003) Detection of single nucleotide polymorphisms. Curr Issues Mol Biol 5:43–60
17. Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E et al (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin Chem 50:1156–1164
18. Makkar HPS, Becker K, Sporer F, Wink M (1997) Studies on nutritive potential and toxic constituents of different provenances of *Jatropha curcas*. J Agric Food Chem 45:3152–3157
19. Makkar HPS, Aderibigde AO, Becker K (1998) Comparative evaluation of non-toxic and toxic varieties of *Jatropha curcas* for

chemical composition, digestibility, protein degradability and toxic factors. Food Chem 62(2):207–215

20. Mittal N, Dubey AK (2009) Microsatellite markers—a new practice of DNA based markers in molecular genetics. Phcog Rev 3:235–246

21. Norambuena PA, Copeland JA, Křenková P, Štambergová A, Macek M Jr (2009) Diagnostic method validation: high resolution melting (HRM) of small amplicons genotyping for the most common variants in the *MTHFR* gene. Clinical Biochem 42(12):308–1316

22. Ovando-Medina I, Espinosa-Garcia FJ, Núnez-Farfán JS, Salvador-Figueroa M (2011) State of the art of genetic diversity research in *Jatropha curcas*. Scientific Res Essays 6:1709–1719

23. Pamidimarri DVNS, Singh S, Mastan SG, Patel J, Reddy MP (2009) 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(6):1357–1364

24. Panigrahi S, Francis BJ, Cano LA, Burbage MB (1984) Toxicity of *Jatropha curcas* seeds from Mexico to rats and mice. Nutr Rep Int 29:1089–1099

25. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Eco Notes 6:288–295

26. Ram SG, Parthiban KT, Senthil KR, Thiruvengadam V, Paramathma M (2008) Genetic diversity among *Jatropha* species as revealed by RAPD markers. Genet Resour Crop Evol 55:803–809

27. Ranade SA, Srivastava AP, Rana TS, Srivastava J, Tuli R (2008) Easy assessment of diversity in *Jatropha curcas* L. plants using two single-primer amplification reaction (SPAR) methods. Biomass and Bioenergy 32:533–540

28. Rohlf FJ (2002) NTSYS-pc. Numerical taxonomy and multivariate analysis system, version 2.10. Exeter Software, New York

29. Rosado TB, Laviola BG, Faria DA, Pappas MR, Bhering LL, Quirino B et al (2010) Molecular markers reveal limited genetic diversity in a large germplasm collection of the biofuel crop *Jatropha curcas* L. in Brazil. Crop Sc 50:2372–2382

30. Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

31. Saha MC, Cooper JD, Mian MAR, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449–1458

32. Sardos J, McKey D, Duval MF, Malapa R, Noyer JL, Lebot V (2008) Evolution of cassava (*Manihot esculenta* Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. Genome 51:912–921

33. Serebryanaya A, Shipunov A (2009) Morphological variation of plants on the uprising islands of northern Russia. Ann Bot Fennici 46:81–89

34. Shabanimofrad M, Yusop MR, Saad MS, Wahab PEM, Biabanikhanehkahdani A, Latif MA (2011) Diversity of physic nut (*Jatropha curcas*) in Malaysia: application of DIVA—geographic information system and cluster analysis. Australian Journal of Crop Science 5(4):361–368 (2011) ISSN:1835–2707

35. Simko I, Pechenick DA, McHale LK, Truco MJ, Ochoa OE, Michelmore RW et al (2009) Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. BMC Plant Biol 9:135

36. Sujatha M, Reddy TP, Mahasi MJ (2008) Role of biotechnological interventions in the improvement of castor (*Ricinus communis* L.) and *Jatropha curcas* L. Biotechnology Adv 26:424–435

37. Sun Q, Li L, Li Y, Wu G, Ge X (2008) SSR and AFLP markers reveal low genetic diversity in the biofuel plant *Jatropha curcas* in China. Crop Sci 48:1865–1871

38. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

39. Weng Y, Azhaguvel P, Michels GJ Jr, Rudd JC (2007) Cross-species transferability of microsatellite markers from six aphid (Hemiptera: Aphididae) species and their use for evaluating biotypic diversity in two cereal aphids. Insect Mol Biol 16:613–622

40. Xu WW, Sleper DA, Krause GF (1994) Genetic diversity of tall fescue germplasm based on RFLPs. Crop Sci 34:246–252