

Article

## DNA Markers and FCSS Analyses Shed Light on the Genetic Diversity and Reproductive Strategy of *Jatropha curcas* L.

Daria Gigliola Ambrosi <sup>1</sup>, Giulio Galla <sup>1</sup>, Marina Purelli <sup>2</sup>, Tommaso Barbi <sup>2</sup>, Andrea Fabbri <sup>2</sup>, Sergio Lucretti <sup>3</sup>, Timothy Francis Sharbel <sup>4</sup> and Gianni Barcaccia <sup>1,\*</sup>

<sup>1</sup> Laboratory of Plant Genetics and Genomics, College of Agriculture, University of Padova, Campus of Agripolis, Viale dell'Università 16, 35020 Legnaro, Padova, Italy;

E-Mails: dariagigliola.ambrosi@studenti.unipd.it (D.G.A.); giulio.galla@unipd.it (G.G.)

<sup>2</sup> Geneticlab Srl, Via Corte Ferrighi 16/B, 36025 Noventa Vicentina, Vicenza, Italy;

E-Mails: marina.purelli@geneticlab.it (M.P.); tommaso.barbi@geneticlab.it (T.B.); andrea.fabbri@geneticlab.it (A.F.)

<sup>3</sup> ENEA, CR Casaccia, BIOTEC Section of Plant Genetics and Genomics, Via Anguillarese 301, 00123 Roma, Italy; E-Mail: sergio.lucretti@enea.it

<sup>4</sup> Apomixis Research Group, Leibniz Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany; E-Mail: sharbel@ipk-gatersleben.de

\* Author to whom correspondence should be addressed; E-Mail: gianni.barcaccia@unipd.it.

Received: 1 March 2010; in revised form: 5 April 2010 / Accepted: 7 May 2010 /

Published: 25 May 2010

---

**Abstract:** *Jatropha curcas* L. ( $2n = 2x = 22$ ) is becoming a popular non-food oleaginous crop in several developed countries due to its proposed value in the biopharmaceutical industry. Despite the potentials of its oil-rich seeds as a renewable source of biodiesel and an interest in large-scale cultivation, relatively little is known with respect to plant reproduction strategies and population dynamics. Here, genomic DNA markers and FCSS analyses were performed to gain insights into ploidy variation and heterozygosity levels of multiple accessions, and genomic relationships among commercial varieties of *Jatropha* grown in different geographical areas. The determination of ploidy and the differentiation of either pseudogamous or autonomous apomixis from sexuality were based on the seed DNA contents of embryo and endosperm. The presence of only a high 2C embryo peak and a smaller 3C endosperm peak (ratio 2:3) is consistent with an obligate sexual reproductive system. Because of the lack of either 4C or 5C endosperm DNA estimates, the occurrence of gametophytic

apomixis seems unlikely in this species but adventitious embryony cannot be ruled out. The investigation of genetic variation within and between cultivated populations was carried out using dominant RAPD and Inter-SSR markers, and codominant SSR markers. Nei's genetic diversity, corresponding to the expected heterozygosity, was equal to  $H_e = 0.3491$  and the fixation index as low as  $F_{st} = 0.2042$ . The main finding is that seeds commercialized worldwide include a few closely related genotypes, which are not representative of the original Mexican gene pool, revealing high degrees of homozygosity for single varieties and very low genetic diversity between varieties.

**Keywords:** *Jatropha* germplasm; seed ploidy; heterozygosity; commercial varieties

---

## 1. Introduction

*Jatropha curcas* L. is a mainly diploid ( $2n = 22$ ) tropical shrub plant belonging to the Euphorbiaceae family, with a relatively small genome of approximately 370 Mb [1]. The species is native to Central and South America, but is now widely present throughout Central America, Africa and Asia. It is becoming a popular non-food oleaginous crop in several developed countries for its proposed value in the biopharmaceutical and biopesticides industry [2]. The short gestation period, easy adaptation to different kinds of marginal lands, even under harsh climatic conditions, relatively high drought- and pest-tolerance, and avoidance by herbivores, make this plant species attractive for cultivation. Its seeds contain up to 40% oil that can be converted into biodiesel by transesterification [3-6]. Nevertheless, it is worth noting that advanced agronomic practices seem to be necessary under a plantation crop regime, being water inputs, fertilizers and pesticides crucial to maximize seed yields [7].

Despite the potentials of its oil-rich seeds as a renewable source of fuel for diesel engines and the interest towards large-scale plantation systems, basic biological information is scanty [8]. For instance, reproduction strategies and population dynamics are poorly understood. The extent of genetic variation within local ecotypes, as well as genetic differentiation between commercial varieties, remains uncharacterized [9,10]. On the basis of available information, the vast majority of seeds should be set through sexuality, mainly by outcrossing, also even though selfing. In addition, apomixis (*i.e.*, asexual propagation through seeds [11]) has also been reported, but is not well documented [12,13].

In the last years, a few genetic characterization studies using molecular markers have been conducted for *J. curcas* germplasm (for reference details, see recently published papers [9,14,15]). Some earlier investigations revealed relatively high estimates of polymorphism either within or between populations using RAPD and Inter-SSR markers alone [16,17] or in combination with SSR markers [18]. By contrast, an exceptionally broad genetic similarity along with phenotypic diversity in *J. curcas* germplasm was found using RAPD and AFLP markers [14]. More recently, a tight clustering of *J. curcas* worldwide accessions, with the exception of Mexican ecotypes, was reported on the basis of RAPD and Inter-SSR markers [9].

The most important finding remains the occurrence of high genetic diversity within autochthonous germplasm from Central America (e.g., Mexico) and low genetic polymorphism between accessions derived from cultivated plantations of African and Asian regions, both in terms of phorbol ester levels and molecular genetic fingerprints [9]. Low genetic variation in a crop typically implies a population bottleneck or a high level of human-mediated selection, but it may also reveal lack of evolution in the corresponding species (for review, see [19]). An evolutionary event like a genetic bottleneck may have prevented from reproducing a significant proportion of *J. curcas* populations, leading over time to a gain of homogeneity. In fact, population bottlenecks are also known to markedly accentuate inbreeding at the species level due to the narrowed pool of genes and reduced combination of genotypes. Moreover, recurrent morphological selection programs may have decreased genetic diversity in *J. curcas* beyond that caused by bottlenecks [19]. As a matter of fact, *J. curcas* has never been bred for productivity and/or quality traits, and all commercial varieties were derived via phenotype-based selections from local germplasm and propagated by vegetative cuttings. As a consequence, limited genetic variation of *J. curcas* populations outside the center of origin is likely, due to infrequent introductions without further augmentation of the crop, predominant vegetative propagation or occurrence of asexual reproduction (*i.e.*, apomixis).

Despite the pronounced phenotypic variability shown by plants and seeds among accessions of different geographical origin [20], a number of contrasting results in terms of heterozygosity have been reported, ranging from very high [18] to moderately low [9], as estimated on the basis of co-dominant SSR markers. Additional studies are thus required to shed light on DNA polymorphism levels both within local ecotypes and between commercial varieties; crucial information for genetic conservation and selection programs in this species. In addition, sub-optimal agronomic practices, along with a low number of cloned genes and its uncharacterized genome, make *J. curcas* a species where major research initiatives in agronomy and biotechnology are required with the aim of breeding new genetically improved varieties [21]. For instance, the adoption of transgenesis for the improvement of biofuel crops, including *J. curcas*, was recently recommended [22], while the exploitation of interspecific crosses among closely related *Jatropha* species was postulated as a strategy for the development of new varieties [14].

Traditional methods based on morphology for understanding genetic variation within and among accessions are largely unsuccessful in *J. curcas* due to the strong influence of agronomic factors and environmental conditions on seed and plant traits. This study deals with the combined use of genomic DNA markers and flow cytometric seed screen (FCSS [23]) analyses to gain new insights into the population genetics and reproduction systems of *Jatropha* spp. Particular attention was given to the evaluation of ploidy levels and heterozygosity degrees of single accessions, and genomic relationships among commercial varieties grown in different countries and continents, and the Mexican landraces by means of co-dominant SSR marker analysis.

## 2. Materials and Methods

### 2.1. Plant Materials

Most of the commercial varieties of *J. curcas* investigated in the present work were purchased from private companies (Table 1).

About 2,500 seeds belonging to a total of 17 varieties, known to be cultivated in different geographical areas, were gathered (Table 1). Concerning the geographical distribution of accession groups, five distinct varieties were collected in Africa (Togo, Mali and Tanzania), nine varieties were obtained from Eastern Asia (India, Sri Lanka) and three varieties were collected in Southern America (Brazil and Argentina). Central America (Mexico) was represented by nine accessions belonging to two ecotypes locally dominant on the Pacific Ocean coast of Jalisco State (Table 1).

**Table 1.** The main information on *J. curcas* germplasm characterized by means of flow cytometric and molecular marker analyses.

Accession Id.	Geographical origin	Variety	Seeds	Plants
JcCLO	America/Mexico	Cruz de Loredo <sup>1</sup>	100	4
JcHDE	America/Mexico	Playon de Mismaloya <sup>1</sup>	10	5
JcBRA	America/Brazil	n.s. <sup>2</sup>	30	5
JcSAM	America/Brazil	n.s. <sup>2</sup>	10	1
JcLOL	America/South America	Lola	200	20
JcTOG	Africa/Togo	n.s. <sup>2</sup>	29	7
JcENG	Africa/Tanzania	n.s. <sup>2</sup>	150	27
JcBER	Africa/Western Africa	Berenice	200	22
JcESM	Africa/Western Africa	Esmeralda	200	18
JcKAR	Africa/Western Africa	Karima	200	10
JcSLK	Asia/Sri Lanka	Anil	230	20
JcPET	Asia/South-East Asia	Petra	200	7
JcSEL	Asia/South-East Asia	Selene	200	19
JcJOR <sup>3</sup>	Asia/Jordan	n.s. <sup>2</sup>	100	5
JcCHH	Asia/India (Chhattisgarh)	Chhattisgarh	200	10
JcUDA	Asia/India (Radjesthan)	Udaipur	60	5
JcKAK	Asia/India (Andhra Pradesh)	Kakinada	200	4
JcTUT	Asia/India (Tamilnadu)	Tuticorin	60	5
JcCOI	Asia/India (Tamilnadu )	Coimbatore	200	10
Overall			2,579	204

<sup>1</sup> Refers to the name of the Mexican locality where the landrace was collected;

<sup>2</sup> The name of the commercial variety was not specified by the private company that provided seeds;

<sup>3</sup> Accessions not investigated by means of SSR markers but only for ploidy by means of FSSC analysis.

Each of the 19 *J. curcas* germplasm resources was represented by a minimum of one and up to 27 accessions (Table 1). All the 204 plants belonging to the *J. curcas* accession groups were investigated by means of RAPD and Inter-SSR fingerprints. With only one exception, three plants per *J. curcas* accession group were then used for SSR marker analysis, for a total of 55 plants. Individual DNA samples from *J. platyphylla*, *J. standleyi* and *J. malacophylla* plants, recently collected in the Pacific Coast of Jalisco State (Mexico), were used as outgroup reference standards.

## 2.2. Ploidy Estimation and Flow Cytometric Seed Screen (FCSS) Analysis of Reproductive Modes

A total of 190 seeds, with a sample size ranging from 1 to 25 per accession group, were used for assessing ploidy and inferring reproductive modes by means of flow cytometry.

Nuclei were isolated from 20–30 mg cotyledonary leaf tissue and 10–20 mg of endosperm by chopping with a sharp scalpel in a glass Petri dish containing 600  $\mu$ L LB01 lysis buffer (15 mM Tris, 2 mM Na<sub>2</sub>EDTA, 80 mM KCl, 20 mM NaCl, 0.5 mM spermine, 15 mM mercaptoethanol, 0.1% Triton X-100, pH 7.4) supplemented with 100 mM citric acid [24]. Nuclei suspensions were filtered through a 36  $\mu$ m nylon mesh and immediately stained with 2  $\mu$ g/mL 4',6-Diamidino-2-phenylindole (DAPI) for ploidy evaluation.

Ploidy evaluation was performed by flow cytometry (FCSS) using a Fluorescence Activated Cell Sorter FACS Vantage SE (Becton Dickinson, La Jolla, CA, USA) equipped with an argon ion laser emitting 200 mW at a wavelength of 351–363 nm. DAPI-stained nuclei were analyzed at a flow rate of 100 nuclei/sec using a 400 nm long pass filter and chicken red blood cell (CRBC) nuclei as external standard for instrumental calibration and fluorescence reference during measurements, CRBC 2C DNA content = 2.15 pg [25]. DNA fluorescence was acquired as maximum fluorescence intensity (pulse height) and time-related parameters (fluorescence pulse area and width). From three up to 25 independent experiments were carried out for each sample and data were analyzed by CellQuest Pro 4.2.1 acquisition and analysis software (Becton Dickinson, La Jolla, CA, USA).

## 2.3. Genomic DNA Isolation and DNA Marker Detection

Genomic DNA was isolated from 100 mg of fresh leaf tissue using the DNeasy Plant mini-kit (QIAGEN) following the recommendations of the manufacturer. The DNA pellets were washed twice with 70% ethanol, dried and resuspended in 100  $\mu$ L of TE 0.1  $\times$  Buffer (Tris-HCl 100 mM, EDTA 0.1 mM pH 8). The quality of DNA samples was assessed by electrophoresis on 0.8% (p/v) agarose gels, and its concentration was determined by optical density reading (DU650 spectrophotometer, Beckman) at 260 nm (1 O.D. = 50  $\mu$ g/mL). The purity was calculated by the O.D.260/O.D.280 ratio and by O.D.210-O.D.310 pattern [26].

Preliminary analyses defined the optimum PCR conditions for molecular marker survey to select the most efficient primer (RAPD, ISSR) or primer combinations (SSR) to reveal molecular polymorphisms in *J. curcas* accessions.

For the detection of RAPD markers, PCR reactions were performed in a 25  $\mu$ L total volume, including 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M each of dCTP, dGTP, dATP and dTTP, 1.5 U of *Taq* DNA polymerase (GE Healthcare), 1  $\mu$ M of a single primer and 20 ng genomic DNA [27]. Ten 10-mer primers from Operon Technologies Inc. were selected on the basis of the number and size of polymorphic bands generated, and reproducibility of banding patterns (Table 2). PCR reactions were performed in a 9700 Thermal Cycler (Applied Biosystems): the temperature profile consisted of an initial denaturation step of 5 min at 95 °C followed by 3 cycles of 2 min at 95 °C, 1 min at 35 °C, 2 min at 72 °C, then 35 cycles of 1 min at 95 °C, 1 min at 36 °C, 2 min at 72 °C, and a final step of 7 min at 72 °C.

**Table 2.** List of primers used to detect RAPD, Inter-SSR, and SSR molecular markers.

Primer name	Sequence (5'-3')	Basic information		
		Ta (°C)	No. amplicons	Size range (bp)
RAPD	10-mer primer	Ta (°C)	No. amplicons	Size range (bp)
OP-A1	CAGGCCCTTC	36	14	300–2,000
OP-A13	CAGCACCCAC	36	16	100–1,500
OP-E20	AACGGTGACC	36	9	300–2,000
OP-G14	GGATGAGACC	36	8	200–2,100
OP-M3	GGGGGATGAG	36	16	250–1,500
OP-M13	GGTGGTCAAC	36	11	200–1,700
OP-P2	TCGGCACGAC	36	16	170–2,500
OP-P6	GTGGGCTGAC	36	13	200–2,000
OP-Q2	TCTGTCCGGTC	36	18	220–2,000
OP-R8	CCCGTTGCCT	36	11	200–1,650
Inter-SSR	SSR-anchored primer	Ta (°C)	No. amplicons	Size range (bp)
I-4	(CA) <sub>8</sub> AC	58	9	300–1,200
I-15	GTGC(AC) <sub>7</sub>	60	12	450–3,000
I-19	CCTGC(AC) <sub>7</sub>	61	10	400–2,000
I-30	(GT) <sub>6</sub> TG	49	9	450–2,000
I-32	(AGC) <sub>4</sub> C	59	6	400–2,100
I-35	(AGC) <sub>4</sub> GA	59	10	480–2,800
SSR	Forward and Reverse primers	Ta (°C)	Repeat	Expected size (bp)
	GCAAGAGCTCCACAGTAGAAGA			
JcSSR1	CTACCAACTCTTTCCAGTCC GCGCTAGAGATGCTGTTTTCT	55	(AG) <sub>15</sub>	168–187
JcSSR2	CACCACTCTGCAAATAGAAC GACAATGTTAATCGAAGGAG	54	(CT) <sub>13</sub>	123–138
JcSSR3	GACGGGAAAATATATGGCTA GTTTTTCAGGATTGAATGCTCT	55	(GA) <sub>24</sub>	164–195
JcSSR4	CGCAATCCCATTTTTATTAT GATGGGCTCTTTGTAGCTTTT	55	(AG) <sub>20</sub>	178–198
JcSSR5	TTCATGCGACTAGAGCCTAC GTCTATTGAGTGGTTGCATTG	56	(CT) <sub>8</sub>	90
JcSSR6	GCTTTCAATATATCATACTGGT TTATCAACCTATCATCCTCAT	55	(TG) <sub>23</sub>	170–182
JcSSR7	GCACAGAATCAAACCTCAAG ATTTAGCAGAACCCCAAGAAC	53	(TG) <sub>18</sub>	148–202
JcSSR8	ATGTCTCTTTTCCATGTCCAA TCAACCCACCCTCATATAAACC	58	(AC) <sub>17</sub>	154–182
JcSSR9	TACACACTTGCCCTGATTTCTG TTCAGCCAGAATCTCAACAGT	60	(AC) <sub>10</sub>	115
JcSSR10	AGGAAGGATGGGAAGTGGGA	64	(CT) <sub>16</sub>	178–184

Inter-microsatellite markers were assayed by using six different I-SSR primers (synthesized by Life Technologies, Inc.) anchored at 3' or 5' terminus of the simple dinucleotide repeat and extended into the flanking sequence by one to four nucleotide residues (Table 2). The PCR protocol for the detection of Inter-SSR markers resembled one previously described for other species [28]: the reaction constituents were 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 150 μM each of dNTPs, 1.5 U of *Taq* DNA polymerase (GE Healthcare), 1 μM of a single primer and 30 ng of genomic DNA, in a 25 μL total volume. PCR reactions were performed in a 9700 Thermal Cycler (Applied Biosystems) under cycling conditions resembling a touchdown profile: an initial denaturation step of 5 min at 95 °C was followed by 7 cycles of 30 second at 95 °C, 30 second at 59 °C and 30 second at 72 °C. Then, the annealing temperature was reduced by 0.7 °C every cycle until a final annealing temperature of 54 °C was reached. The last cycle was repeated 26 times and terminated by a final step at 72 °C for 10 min.

Both RAPD and Inter-SSR amplicons were separated by electrophoresis in 2% agarose gels run with 1 × TBE buffer [26]. Dominant DNA fingerprints were evaluated by the 1D Image analysis software (Kodak) in order to find amplification products within single lanes and to reliably score qualitative polymorphisms among accessions for all primers.

Microsatellite (SSR) loci analysis was carried out following an already tested PCR protocol [29] with some changes to adapt it to *J. curcas* templates with the use of the 5' M13-tailed primer method [30]. DNA fragments were visualized by capillary electrophoresis after amplification reactions performed with the universal M13 primer labeled with a 5' HEX fluorophore. PCR experiments were conducted in a 20 μL total volume, including 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 200 mM of each dNTP, 3 pmol of primer forward, 8 pmol of primer reverse, 6 pmol M13-labeled primer, 1 U *Taq* DNA polymerase (GE Healthcare) and 25 ng of genomic DNA as template. A total of 10 SSR loci were investigated in all genomic DNA samples. Six of the SSR loci (*i.e.*, JcSSR1-JcSSR6) were assayed using primer pairs derived from a previously reported study [31], whereas the remaining four primer pairs were designed on SSR-containing sequences retrieved from the NCBI database (accession numbers EU099518; EU099528; EU099520; and EU099522). All primer combinations are reported in Table 2. Amplification reactions were performed in a 9700 Thermal Cycler (Applied Biosystems): the temperature profile consisted of an initial denaturation step of 5 min at 95 °C followed by 40 cycles of 30 second at 95 °C, 30 second at annealing temperature of 55–64 °C, and 30 second at 72 °C, followed in turn by 7 min at 72 °C and then held at 4 °C.

DNA fragment analysis was carried out using a fully automated capillary electrophoresis system (Applied Biosystems 3130) and SSR patterns were visualized and scored using the software PEAK SCANNER software version 1.0 (Applied Biosystems).

#### 2.4. Genetic Diversity and Differentiation Statistics

Measures of genetic variability were used to estimate the levels of polymorphism within and between different *J. curcas* accessions.

For SSR markers, the average number of alleles observed per locus ( $n_o$ ) was computed as the arithmetic mean over loci of the total number of alleles observed at each locus. The effective number of alleles per locus was computed as:  $n_e = 1/\sum p_i^2$  where  $p_i$  is the frequency of the  $i^{\text{th}}$  allele [32]. This

parameter is a measure of diversity and indicates the size of an ideal population in which, given the existing allele frequencies, all individuals are totally differentiated.

Genetic diversity statistics were used to summarize the data of the *J. curcas* SSR markers. For each marker locus and over all loci, the genetic diversity was computed as  $H = 1 - \sum p_i^2$  corresponding to the Nei's expected heterozygosity [33]. This parameter ranges from 0 in case of monomorphic allele in all loci to 1 when alleles in equal frequencies occur over all polymorphic loci. Moreover, the phenotypic diversity of marker allele profiles was also estimated using the Shannon's information index, computed as  $I = -\sum p_i^2 \ln p_i^2$  [34].

Gene flow estimates were derived from the fixation index as follows:  $Nm = 0.25(1 - F_{st})/F_{st}$ . The result is independent of the population size because the force of gene flow, which is measured by the fraction of migrants in a population (denoted as  $m$ ), is counteracted by the force of genetic drift, which is proportional to the inverse of the population size ( $N$ ).  $Nm < 1$  indicates a local differentiation of populations, while  $Nm > 1$  when a little differentiation among populations occurs [35].

A hierarchical analysis of variance with estimation of F-statistics [36] was also performed. Estimates of heterozygosity within accession groups ( $F_{is}$ ) and between accessions ( $F_{it}$ ) were determined as well as the fixation index ( $F_{st}$ ) according to Wright's statistics [37]. The inbreeding coefficient  $F_{is}$  was computed also for single SSR loci to measure the deficiency (+values) or excess (-values) of heterozygotes at each locus and in all accession groups. The  $F_{st}$  parameter measures the genetic effect of total population subdivision as the proportional reduction in overall heterozygosity owing to variation in SSR allele frequencies among different accession groups. The Slatkin's  $R_{st}$  [38] was also computed as it provides a convenient approach for estimating levels of genetic differentiation between populations on the basis of variance for marker allele frequencies. This coefficient, adapted to SSR loci by assuming a high-rate stepwise mutation model instead of a low-rate or infinite-allele mutation model, was calculated as follows:  $R_{st} = (S_t - S_w)/S_t$ , where  $S_w$  is the sum over all loci of twice the weighted mean of the within-population variances, and  $S_t$  is the sum over all loci of twice the variance of the combined populations [39].

Descriptive genetic diversity and differentiation statistics, as well as inbreeding coefficients at each locus and over all loci for each accession group and over all accessions, were conducted using the software POPGENE version 1.21 [40].

An ordination analysis was performed according to the unweighted pair-group arithmetic average method (UPGMA) clustering algorithm [41], and the dendrogram and centroids of all accessions were constructed from the symmetrical genetic similarity matrix. The simple matching (SM) coefficient was applied to calculate the proportion of genetic similarity between all pair-wise comparisons of accessions. The following formula was used:  $SM_{ij} = (p + n)/t$ , where  $p$  and  $t$  represent the number of marker alleles, respectively, present and absent in the pair of individuals  $i$  and  $j$  compared, and  $t$  is the total number of marker alleles scored in the population as a whole. The principal coordinate analysis technique [42] was then applied to compute the first two components out of the qualitative data matrix. The triangular matrix of genetic similarity estimates was double-centered and then bi-dimensionally plotted according to the extracted Eigen-vectors. The calculations and analyses were conducted using the appropriate routines of the software NTSYS version 1.80 [43].

The population structure was investigated using the Bayesian (model-based) clustering algorithm implemented in the STRUCTURE software [44,45], which identifies subgroups of individuals

according to marker allele combination and distribution. This software was also exploited to assign each individual DNA sample of varieties and landraces, pre-defined according to the geographical origin, to the inferred clusters. All simulations were carried out assuming the admixture model, with no a priori population information. Analyses of SSR data were carried out with 500,000 iterations and 500,000 burn-ins, by assuming the allele frequencies among populations to be correlated [45]. Ten replicate runs were performed with each run, exploring a range of K spanning from 1 up to 16. Estimation of the most likely value of K was done using  $\Delta K$  as reported in other studies [46].

### 3. Results

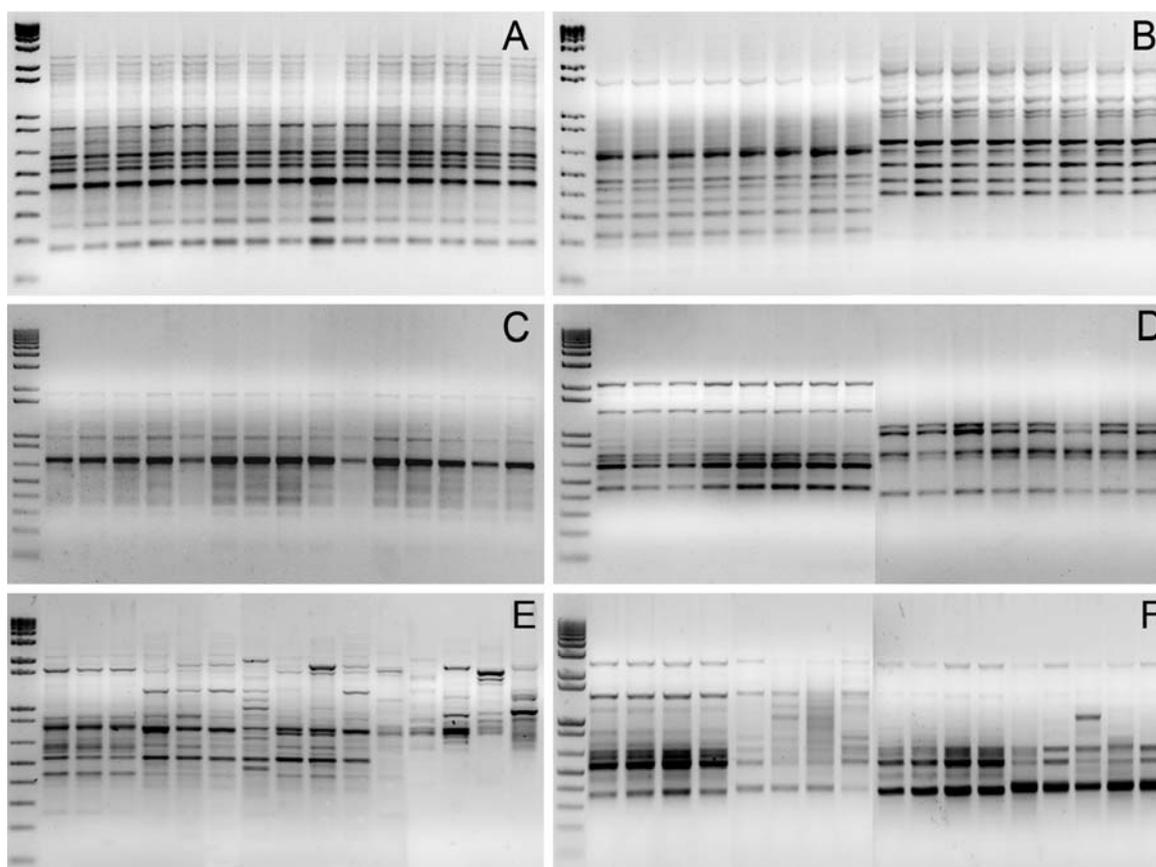
#### 3.1. Genomic DNA Fingerprint Analysis By Means Of RAPD and Inter-SSR Markers

Genetic diversity and differentiation among *J. curcas* germplasm resources from different geographical areas were initially investigated using dominant RAPD and Inter-SSR markers, with primers selected during preliminary experiments on the basis of the total number of reproducible amplicons. A total of ten 10-mer primers and six SSR-anchored primers (see Table 2) were used to assay 142 and 61 genomic loci, respectively, with an average number of markers per DNA fingerprint equal to 13 and 9. Surprisingly, low genetic variability was documented not only within accession groups but also among accessions of different geographical origin, with the exception of Mexican landraces (Figure 1). In fact, no reliable polymorphism was detected within or between commercial varieties: RAPD and Inter-SSR fingerprints visualized by conventional agarose gel electrophoresis were 100% monomorphic (see Figure 1, panels A-D). As few as nine and five polymorphic loci out of 203 assayed in total (6.9%) were found among landrace individuals using RAPD and Inter-SSR markers, respectively, with the most informative primers being OP-P2 and I-19 (see Figure 1, panels E and F).

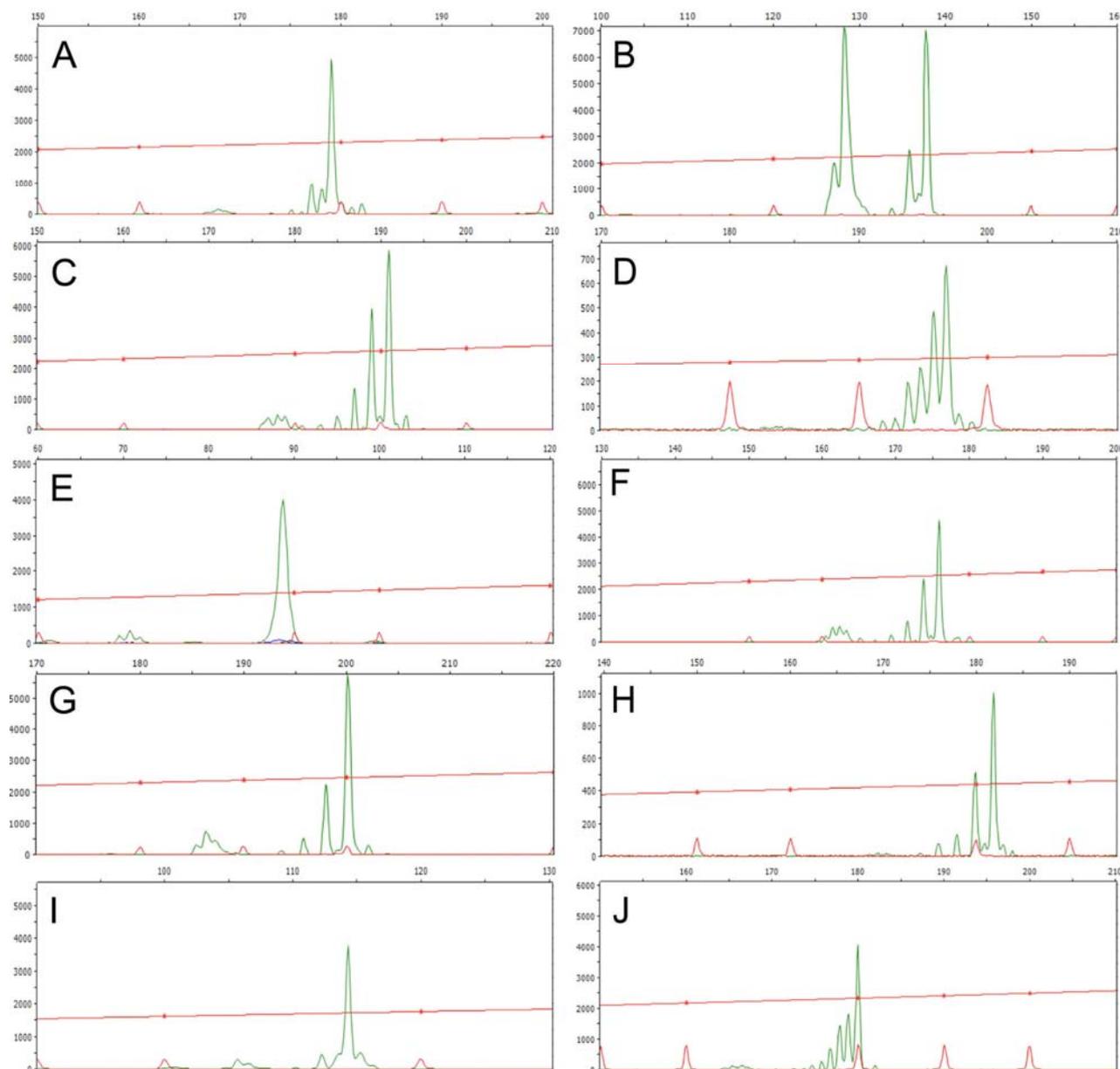
#### 3.2. Genetic Diversity and Differentiation Statistics Based On SSR Markers

Genetic diversity of *J. curcas* accessions was further investigated using SSR markers with primer pairs designed in order to maximize their specificity and PCR reproducibility. A total of 10 microsatellite loci (see Table 2) were assayed using one to four individual DNA samples per locality/variety, for a total number of 53 plants belonging to American (7), African (15), Asian (22) varietal groups and Mexican landrace accessions (9). Moreover, one DNA sample each of *J. malacophylla*, *J. standleyi* and *J. plathyphylla*, adopted as outgroup standards, was successfully analyzed using heterologous primers. An overview of SSR markers as detected by DNA fragment analysis with a fully automated capillary electrophoresis system is given in Figure 2.

**Figure 1.** Genomic DNA fingerprints of *J. curcas* accessions. **A, B:** Examples of RAPD fingerprints generated using primer OP-P2 (**A**), and OP-M13 and OP-A1 (**B**): no polymorphism is detectable either within a population (**A**) nor between populations (**B**) of different geographical origin (**A**, South-East Asia; **B**, England vs. South-America and Sri Lanka vs. England). **C, D:** Inter-SSR fingerprints generated using the primer I-15 with genomic DNA templates of three individuals for each of the accession groups: JcSLK, JcBER, JcLOL, JcESM, JcENG (**C**). Inter-SSR fingerprints generated using primers I-4 (left) and I-19 (right) with genomic DNA templates of four individuals for each of the accession groups JcTOG and JcBRA (**D**). **E, F:** RAPD fingerprints generated using primer OP-P2 in three commercial varieties, nine Mexican accessions and three outgroups (**E**); Inter-SSR fingerprints generated with primer I-15 (left) and I-19 (right) using DNA templates belonging to four varieties and four Mexican accessions (**F**).



**Figure 2.** Representative DNA fragment profiles displayed by each of the 10 analyzed microsatellite loci. **A**, JcSSR1; **B**, JcSSR2; **C**, JcSSR3; **D**, JcSSR4; **E**, JcSSR5; **F**, JcSSR6; **G**, JcSSR7; **H**, JcSSR8; **I**, JcSSR9; and **J**, JcSSR10 (for details on primers, see Table 2).



Unexpectedly, the vast majority of these loci showed homozygosity or heterozygosity for the same microsatellite alleles across all accessions from Southern America and Africa: several examples of SSR stuttered patterns for homozygous loci are given in **Appendix 1**.

The Mexican accessions proved to be genetically differentiated from commercial varieties, showing eight polymorphic SSR loci out of 10 and a total of 29 marker alleles. It is worth mentioning that aspecific amplification products were occasionally recovered for some combinations of primer pairs and accession templates (data not shown).

Furthermore, most of the accessions from India revealed heterozygosity for different microsatellite alleles at some of the loci investigated (see **Appendix 2**).

The main descriptive statistics of *J. curcas* germplasm related to SSR markers were determined at single loci and over all accession groups. A total number of 38 marker alleles were detected across the 10 assayed SSR loci, with a mean frequency of the commonest marker allele equal to  $p_i = 0.7372$ , ranging from 0.3962 to 1 (Table 3).

**Table 3.** Descriptive statistics of *J. curcas* germplasm related to SSR markers and accession groups. Sample size of marker alleles (S), frequency of the commonest marker allele ( $p_i$ ), observed number of alleles ( $n_o$ ) and expected number of alleles ( $n_e$ ) per locus, and estimates of Shannon's information index of phenotypic diversity (I) and Nei's genetic diversity equivalent to the expected heterozygosity (H). The Wright's inbreeding coefficient ( $F_{is}$ ) is also given for single locus, geographical regions and over all accessions. The overall values and standard deviations are reported for each parameter.

Accession Id.	S	$p_i$	$n_o$	$n_e$	I	H	$F_{is}$
JcSSR1	106	0.8962	4	1.2372	0.4275	0.1917	0.7047
JcSSR2	106	0.3962	4	3.2474	1.2656	0.6921	-1.036
JcSSR3	106	0.5943	6	2.5069	1.25	0.6011	0.4664
JcSSR4	106	0.5472	6	2.7259	1.2407	0.6331	0.255
JcSSR5	106	1	1	1	0	0	0
JcSSR6	104	0.9231	4	1.1706	0.3512	0.1457	0.736
JcSSR7	98	0.949	3	1.1082	0.227	0.0977	0.791
JcSSR8	100	0.47	4	2.3397	0.9612	0.5726	-0.5718
JcSSR9	104	1	1	1	0	0	0
JcSSR10	104	0.5962	5	2.0969	0.8902	0.5231	-0.3234
Mexico	17	0.6611	2.9	2.0615	0.7485	0.4260	0.1138
South America	14	0.7857	2	1.5949	0.4407	0.2663	-0.1609
Africa	30	0.8333	1.6	1.4241	0.3115	0.1978	-0.4116
Asia	43	0.75	2	1.6789	0.481	0.2999	0.0998
Mean	104	0.7372	3.8	1.8433	0.6613	0.3457	-0.1438
St. Dev.	-	0.2373	1.7512	0.8356	0.5168	0.2819	-

The observed and expected number of alleles per locus varied from 1 to 6.0000 and 3.2474, respectively, being on average equal to  $n_o = 3.8000$  and to  $n_e = 1.8433$  (Table 3). For these parameters, Mexican accessions scored the highest values ( $n_o = 2.9000$  and  $n_e = 2.0615$ ). The average estimate of the Shannon's information index of phenotypic diversity for molecular profiles was  $I = 0.6613$ , whereas the mean Nei's genetic diversity, equivalent to the expected heterozygosity, was as low as  $H = 0.3457$  (Table 3). The estimates of genetic diversity among accessions grouped according to their geographical origin ranged from a minimum of 0.1978 to a maximum of 0.4260. In particular, little genetic variation was found either within or between commercial varieties deriving from different parts of the world, and most of the gene/allele variation and differentiation noticed exist in the landraces from Central America. Moreover, the inbreeding coefficient was on average equal to  $F_{is} = -0.1438$ , with values that suggested a marked deficiency of heterozygosity for varieties from Africa and Southern America, and an excess of heterozygosity for Mexican and Indian materials (Table 3).

Summary of heterozygosity (H), F-statistics and gene flow (Nm) estimates for individual marker loci, accession groups and over all accessions is reported in Table 4. Wright's inbreeding coefficients  $F_{is}$  and  $F_{it}$  were calculated as measures of heterozygosity excess (+) or deficiency (-) of single individuals compared to accession groups and the total population, respectively. In particular, the average value of  $F_{it} = 0.0897$  revealed slight variations of individual heterozygosity degrees from the overall population heterozygosity estimate. The very low amount of genetic differentiation  $F_{st} = 0.2042$ , also known as fixation index, indicates that only 20% of the genetic variation was found among accessions and that as much as 80% of the total polymorphism was scored within accession groups. A moderate genetic differentiation among geographical groups was also supported by a relatively low value of  $R_{st} = 0.2927$ . The narrow genetic differentiation among varietal groups over the investigated SSR loci was confirmed by the gene flow estimate that on average resulted  $Nm = 0.94$  (Table 4).

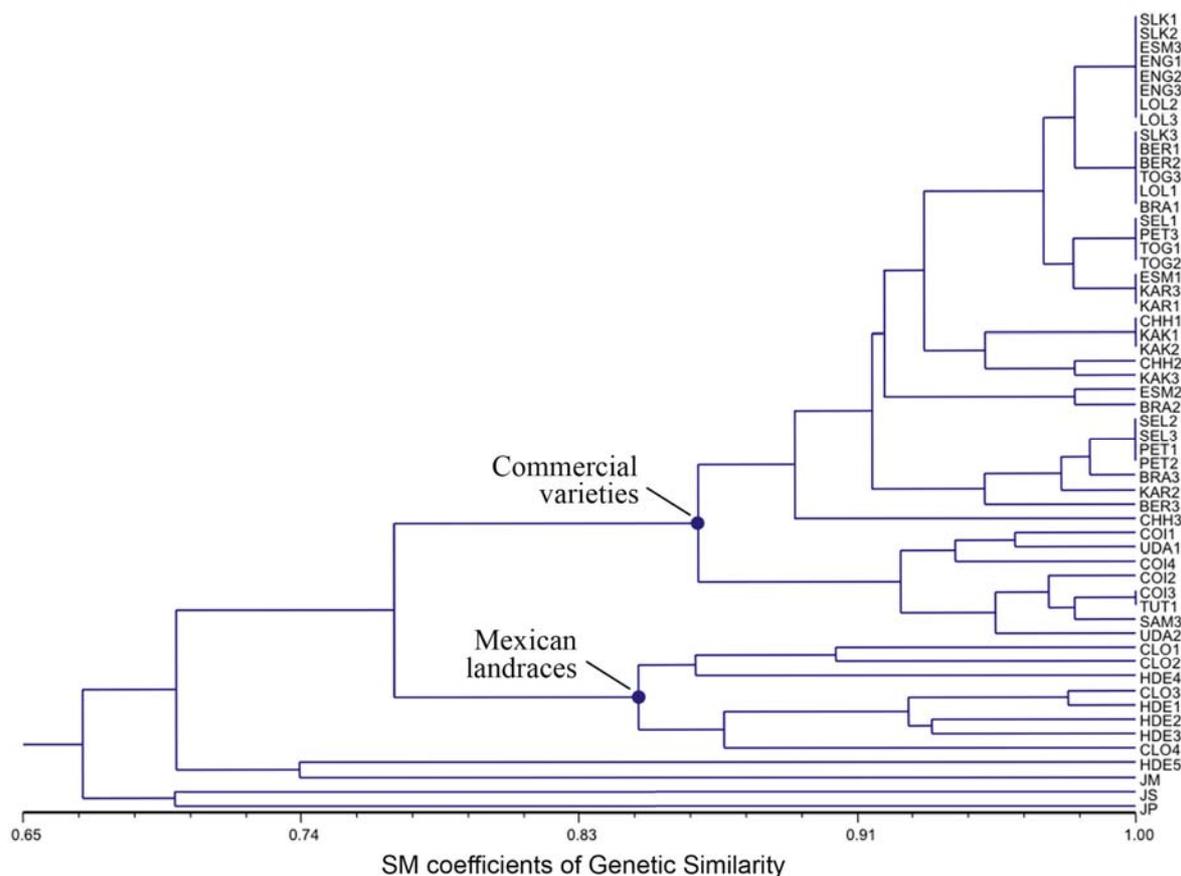
**Table 4.** Summary of genetic diversity values (H), F-statistics and gene flow (Nm) estimates for individual marker loci and over all accessions. Nei's genetic diversity estimates were calculated as observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) and average heterozygosity ( $H_{av}$ ), whereas Wright's inbreeding coefficients  $F_{is}$  and  $F_{it}$  were calculated as measures of heterozygote deficiency (-) or excess (+) of single individuals compared to accession groups and the total population, respectively, while  $F_{st}$  is the degree of genetic differentiation, also known as fixation index.

Marker Locus Id.	Sample Size	H-statistics			F-statistics			Gene flow
		$H_o$	$H_e$	$H_{av}$	$F_{is}$	$F_{it}$	$F_{st}$	Nm
JcSSR1	106	0.0566	0.1936	0.1683	0.5048	0.6582	0.3096	0.56
JcSSR2	106	0.717	0.6987	0.4832	-0.4287	0.0339	0.3238	0.52
JcSSR3	106	0.3208	0.6068	0.3985	0.1659	0.4413	0.3302	0.5
JcSSR4	106	0.4717	0.6392	0.5684	0.067	0.2016	0.1443	1.48
JcSSR5	106	0	0	0	0	0	0	0
JcSSR6	104	0.0385	0.1471	0.1665	0.7172	0.7415	0.0859	2.66
JcSSR7	98	0.0204	0.0987	0.1354	0.6923	0.7808	0.2877	0.62
JcSSR8	100	0.9	0.5784	0.5753	-0.5078	-0.4804	0.0182	13.52
JcSSR9	104	0	0	0	0	0	0	0
JcSSR10	104	0.6923	0.5282	0.4794	-0.5057	-0.2708	0.156	1.35
Mean	104	0.3217	0.3491	0.2975	-0.1438	0.0897	0.2042	0.97
St. Dev.	-	0.3494	0.2846	0.2274	0.4269	0.3567	0.1379	4.08

### 3.3. Ordination Analyses of Accessions and Population Structure Inferences

Cluster analysis was used to construct an UPGMA dendrogram displaying the very high genetic similarity detected among the 44 *J. curcas* accessions belonging to commercial varieties (Figure 3).

**Figure 3.** Dendrogram of *J. curcas* accessions based on SSR marker data using the Simple Matching genetic similarity matrix and the UPGMA clustering method. The two main nodes refer to the subgroups of commercial varieties and Mexican landraces.



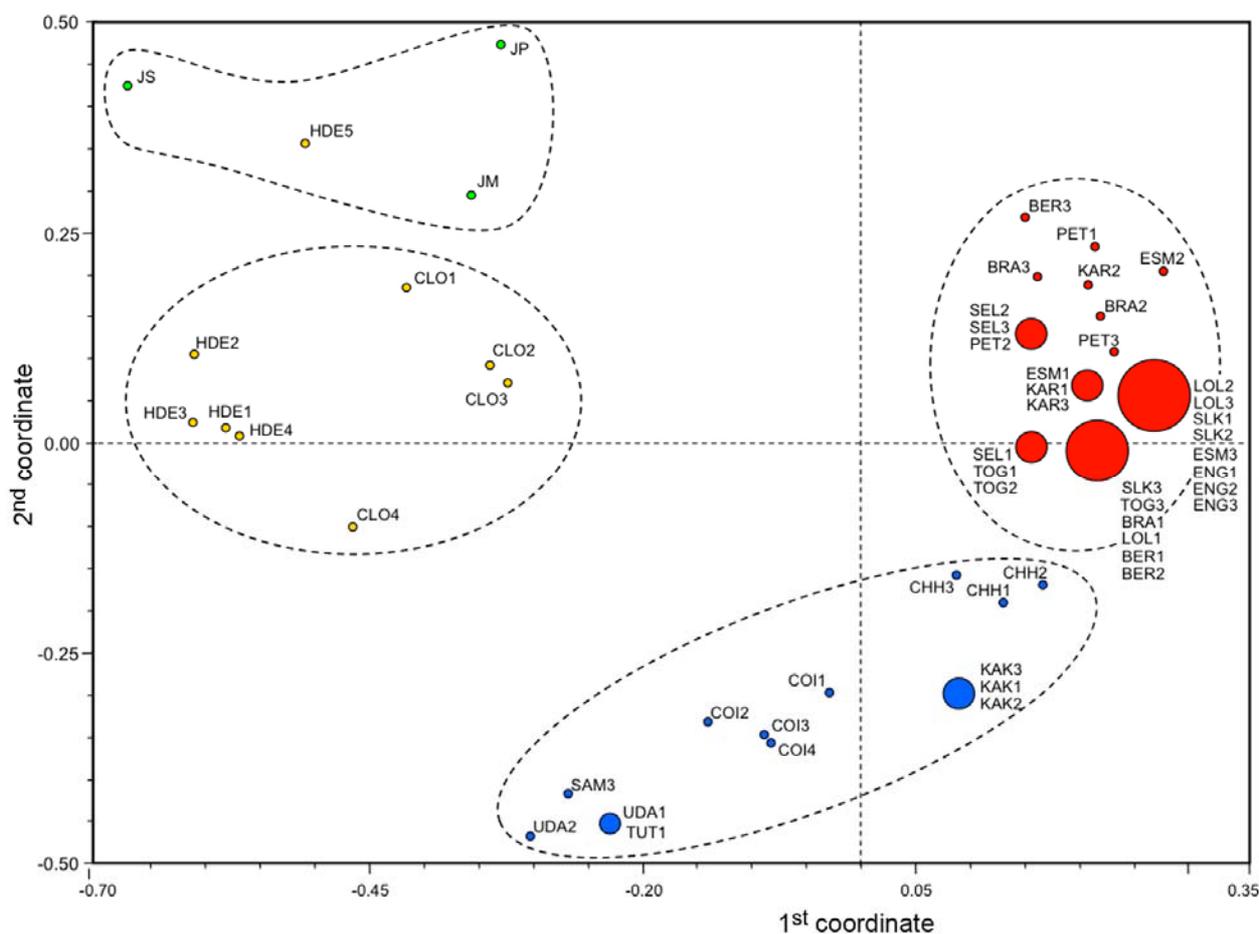
All of them were clustered into two distinct subgroups of 36 and eight accessions each, showing a mean genetic similarity higher than 94%, whereas the remaining accessions that correspond to the eight Mexican landraces were clustered apart with a mean genetic similarity of about 77% (Figure 3). Within the major subgroup, 21 out of the 36 accessions belonging to commercial varieties were tightly clustered into four closely related clonal genotypes, scoring a mean genetic similarity higher than 99%. As expected, the *J. platyphylla*, *J. standleyi* and *J. malacophylla* accessions were grouped separately from *J. curcas* ones. It is worth mentioning that one of the Mexican accessions proved to be genetically differentiated from the remaining *J. curcas* ones, being closely clustered with *J. malacophylla* (73.7%), *J. standleyi* (70.3%) and *J. platyphylla* (63.2%). On the whole, a total of 24 highly correlated genotypes were found over all commercial varieties, with a few genotypically distinct subgroups.

A principal coordinate ordination analysis was performed using the whole set of monomorphic and polymorphic SSR marker alleles according to the UPGMA clustering algorithm [41]. Varieties cultivated in India were clustered separately from the remaining ones collected in Sri Lanka, Western Africa and Southern America that in turn were grouped into a few dominant genotypes (Figure 4).

The first two principle components were able to explain 44.2% of the total genetic variation found within the *Jatropha* germplasm. In particular, the first component, which explains 28.8% of the total genetic diversity, was positively associated with the vast majority of commercial varieties and

negatively associated with Mexican materials. The second component, which explains 15.3% of the total genetic diversity, was clearly able to discriminate the *J. curcas* Mexican landraces from the *J. curcas* Indian varieties, and also from outgroups of *Jatropha* spp., with the exception of the accession coded as HDE5 (Figure 4). The most discriminant microsatellite loci between varieties and landraces proved to be JcSSR1 and JcSSR3.

**Figure 4.** Centroids of *J. curcas* accessions bi-dimensionally plotted according to the principal coordinates that explain 44.2% of the total genetic variation detected across all SSR marker loci. The first component is positively associated with the vast majority of commercial varieties and negatively associated with Mexican materials and outgroups, whereas the second component discriminates the *J. curcas* Mexican landraces from either *J. curcas* Indian varieties and *J. platyphylla*, *J. standleyi* and *J. malacophylla* samples.

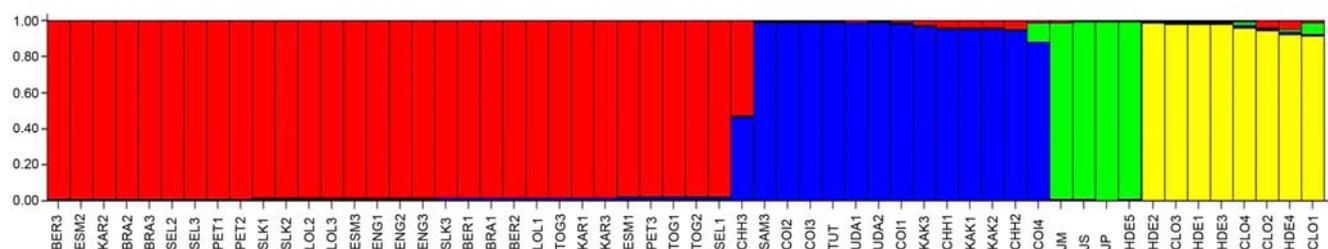


Investigation of the population structure by estimation of  $\Delta K$  [46] suggested that our core collection of accessions is most likely made up of four genetically distinguishable subgroups ( $K = 4$ ), as shown in Figure 5.

A secondary peak at  $K = 2$  was also detected (data not shown), suggesting the presence of two hierarchical levels of genetic structure in the *Jatropha* plant materials. Using  $K = 2$ , two main gene pool clusters were identified with all Mexican landrace accessions grouped separately from all the other varietal accessions. Increasing the number of populations to  $K = 4$  led to the identification of two major clusters corresponding to varieties from Western Africa, Sri Lanka and Southern America, and

India, respectively, and two additional clusters including, on one side, Mexican landraces, and outgroups (*Jatropha* spp.), on the other. Despite the steady genetic differentiation among clusters, a high genetic homogeneity was observed within each subgroup (Figure 5).

**Figure 5.** Population structure of *Jatropha* spp. germplasm into four major clusters (K = 4) as defined according to individual assignment. Main subgroups represent, from left to right, commercial varieties from Southern America, Sri Lanka and Western Africa (red bars), Indian varieties (blue bars), *Jatropha* spp. outgroups (green bars) and Mexican landraces (yellow bars).



It is worth noting that the accession coded as CHH3 proved to be admixed, sharing large fractions of genetic background with the two main subgroups of commercial varieties. Furthermore, the accession coded as SAM3 deriving from Southern America was clustered with Indian accessions (see Figure 5).

Proportions of membership to the four inferred clusters of each of the five *Jatropha* subgroups pre-defined according to their geographical origin are reported in Table 5. Cultivated accessions were grouped in two distinct varietal clusters with a proportion of membership ranging from a minimum of 84.4% to a maximum of 98.3%. Accessions from Mexico were divided into two main clusters, one including landraces and the other outgroups (Table 5).

**Table 5.** Proportion of membership to the four inferred clusters of each of the five *Jatropha* subgroups pre-defined according to their geographical origin.

Geographical origin	Inferred clusters				No. accessions
	1: varieties	2: varieties	3: landraces	4: outgroups	
Mexico (1)	0.012	0.007	<b>0.642</b>	<b>0.339</b>	12
South America (2)	<b>0.844</b>	0.149	0.004	0.004	7
Africa (3)	<b>0.982</b>	0.011	0.004	0.003	15
Asia (4)	<b>0.983</b>	0.01	0.004	0.004	9
India (5)	0.06	<b>0.924</b>	0.004	0.012	13

### 3.4. FCSS Analysis for the Estimation of Seed DNA Contents

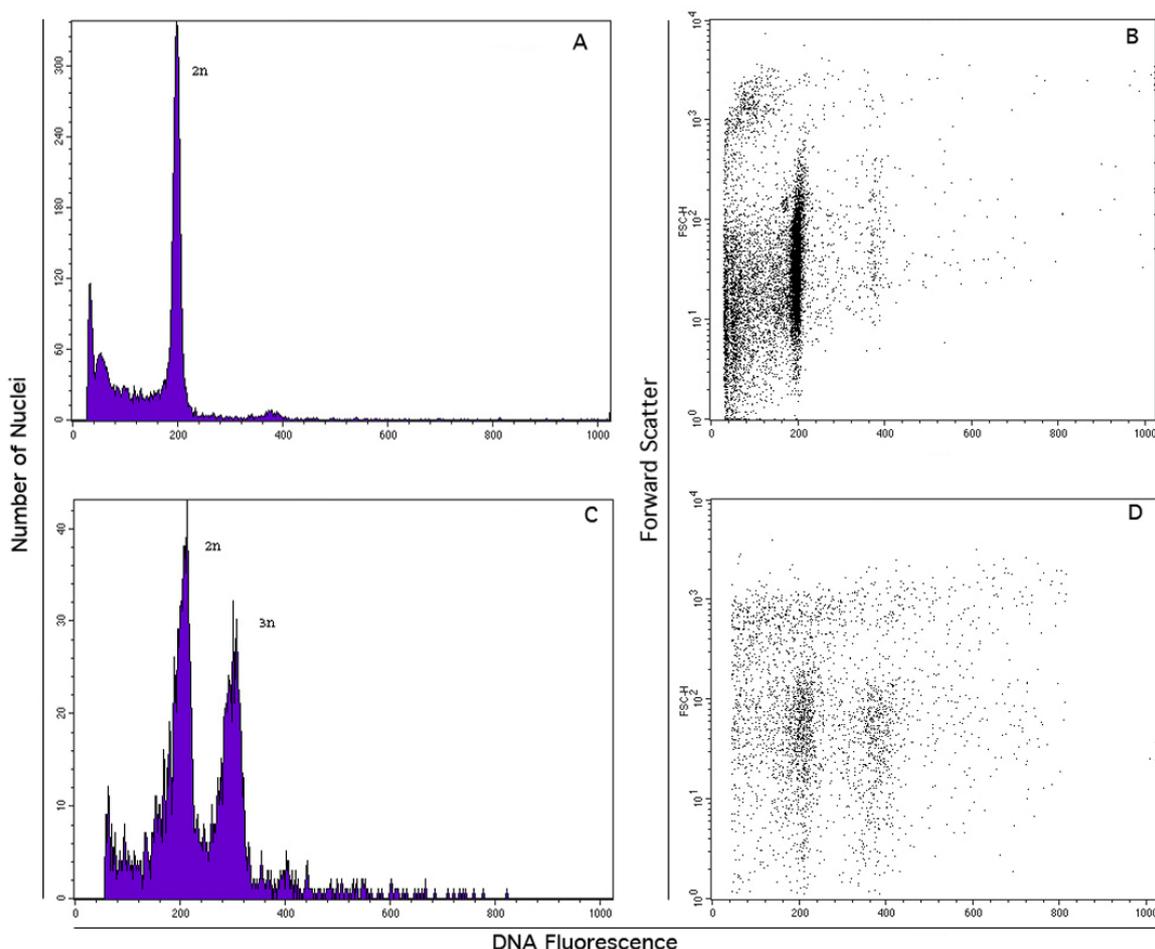
The determination of ploidy and the reconstruction of reproductive strategy in the populations of *J. curcas* were determined by means of FCSS, a method suitable for the discrimination of either pseudogamous or autonomous apomixis from sexuality (*i.e.*, amphimixis) based on the seed DNA contents of embryo and endosperm.

The true DNA content for *J. curcas* has been already estimated as 0.85 pg DNA per 2C nucleus [44], but no extensive analysis of the ploidy level at the population level has been carried until now for plants and seeds.

The high starch content of endosperm tissue required a careful optimization of the chopping procedure [44] and a fine tuning of the relative amount of cotyledon and endosperm tissues, which otherwise could bring wrong data and mistaken ploidy level estimates. Reliable DNA fluorescence pulse area histograms were obtained for the vast majority of analyzed seeds. In particular, the goodness of flow cytometry data was strongly related to the quantity and relative amount of chopped embryo and endosperm tissues. Moreover, the use of LB01 isolation buffer generated high quality DNA histograms when supplemented with 100 mM citric acid. The beneficial effect of citric acid has been already reported as related to a more homogeneous staining of DNA due to a better accessibility of chromatin structure [44,46].

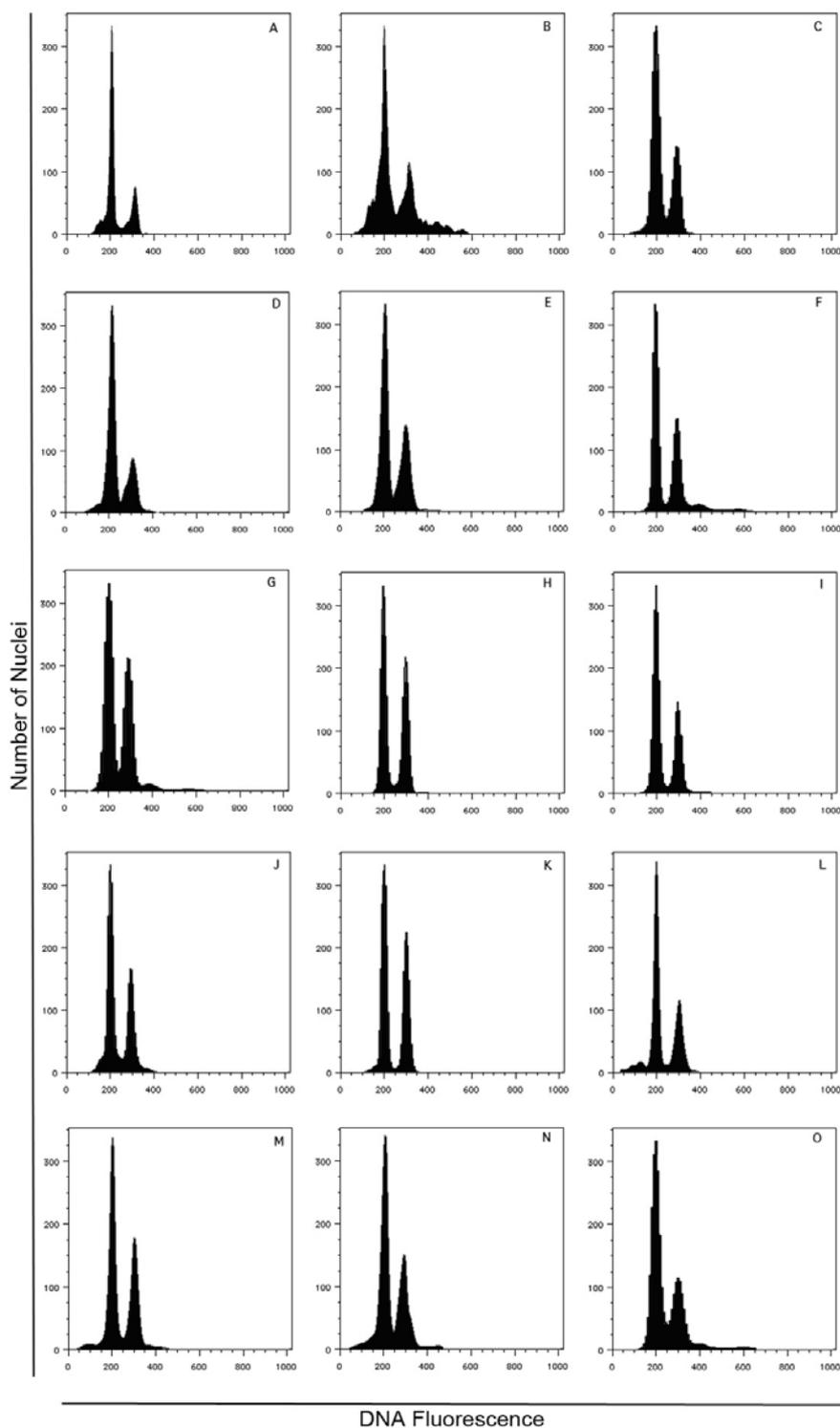
The vast majority of seeds of this species proved to include diploid (2x) embryos and triploid (3x) endosperms (Figure 6).

**Figure 6.** Flow cytometry analysis of DAPI stained *J. curcas* nuclei in suspension. (A) and (C): total DNA fluorescence emission from cotyledonary leaf nuclei alone and together with endosperm released nuclei, respectively; (B) and (D): dual parameter analysis of DNA fluorescence *versus* nuclear scatter light emission for particle discrimination showing cotyledonary nuclei alone and a dual sample with endosperm ones, respectively.



The presence of only one high 2C embryo peak and a smaller 3C endosperm peak (ratio 2:3), without the presence of any other DNA ploidy level, is consistent with an obligate sexual reproductive system (Figure 7).

**Figure 7.** Ploidy evaluation by flow cytometry analysis of DAPI stained nuclei in suspension isolated from single seeds of 15 *J. curcas* accessions: A, JcTUT; B, JcCOI; C, JcUDA; D, JcESM; E, JcPVL; F, JcTAN; G, JcLOL; H, JcBER; I, JcSLK; J, JcTOG; K, JcKAR; L, JcJOR; M, JcBRA; N, JcPET; and O, JcSEL (for details on accessions, see Table 1).



Most importantly, neither tetraploid (4x) nor pentaploid (5x) endosperm DNA estimates were scored, as expected in case of autonomous and pseudogamous apomixis, respectively. However, it is worth mentioning that one triploid was found out of 190 seeds analyzed in this study (Table 6).

**Table 6.** Results of the flow cytometric screening of *J. curcas* seeds. Information on accessions and number of analyzed seeds per population is given along with the number of embryos and endosperm successfully evaluated for the DNA contents of their nuclei.

Accession Id.	No. seeds	2x embryo	3x endosperm
JcCLO	4	4	4
JcHDE	5	5	5
JcBRA	2	1	1
JcSAM	1	1	1
JcLOL	25	21	12
JcTOG	2	2	2
JcENG	20	19	18
JcBER	22	19	8
JcESM	23	12	5
JcKAR	2	1	1
JcSLK	22	21	3
JcPET	22	17	5
JcSEL	22	19	10
JcJOR	2	1	1
JcCHH	2	1	1
JcUDP	2	1	1
JcKAK	5	4	3
JcTUT	4	4	4
JcCOI	3	2	2
	190	155	87

## 4. Discussion

### 4.1. Analysis of Plant DNA Markers and Seed DNA Contents Shed Light on the Genetic Diversity and Reproductive Biology of *J. curcas*

A monoic plant with unisexual flowers, *J. curcas* shows male and female flowers in the same raceme (diclin inflorescences). As far as the life cycle is concerned, it is a perennial rootstock that may be propagated by cuttings or seeds.

On the basis of available information, about 68% of seeds are thought to be set through amphimixis, mainly by outcrossing, even if the species is characterized by self-compatibility and thus selfing is also possible. It was reported that sexual reproduction occurs only through pollination between different flowers of the same plant or from different plants [12]. The species seems also to show a tendency to promote xenogamy (*i.e.*, union of genetically unrelated organisms) and to minimize geitonogamy (*i.e.*, the pollination of a flower with the pollen from another flower on the same flowering plant), mechanisms that increase diversity [12].

That *J. curcas* plants might also set seeds through apomixis was also reported. At the population level, the average degree of apomixis was found to be equal to 32% [11]. It is worth mentioning that agamospermy, as a mode of asexual reproduction through seed, leads to clonality.

The remarkable genetic uniformity within populations is not compatible with outcrossing as the main pollination strategy: in this case, each population would have displayed a heterogeneous mixture of highly heterozygous genotypes sharing a common gene pool. *Vice versa*, the vast majority of microsatellite loci revealed a homozygous state across all investigated plants (the observed homozygosity was equal to 67.8%, on average). Moreover, although low total genetic variation was found, most of that was scored within accession groups (about 80%). As a consequence, *J. curcas* varieties cultivated worldwide nowadays seem to have been bred from closely related germplasm resources.

To establish the reproductive pathway of *J. curcas*, the DNA content of nuclei released from seed embryo and endosperm was measured with flow cytometry. In the case of sexuality, seeds showing diploid 2C embryo combined with triploid 3C endosperm DNA contents (EBN 2m:1p) are expected. By contrast, apomictic reproduction can take place by either pseudogamous (frequent) or autonomous (rare) mechanisms of endosperm development, depending on the occurrence or absence of the central cell fertilization [22]. As a consequence, DNA contents of endosperm nuclei may be either tetraploid (EBN 4m:0p) or pentaploid (EBN 4m:1p) (for details, see [22]). Gametophytic apomixis was not found because neither 4C (autogamous) nor 5C (pseudogamous) endosperm DNA values were scored, as instead it was expected in case of apomictically formed seeds [47,48]. As a consequence, the occurrence of asexual reproduction by means of gametophytic apomixis seems unlikely in this diploid species. Additionally, gametophytic apomixis has been shown to be strongly correlated with the occurrence of hybridity and polyploidy. Although numerous apomeiotic mutants and parthenogenic mutants have been described in diploid forms of sexual species, the expression of gametophytic apomixis is restricted mostly to polyploid, highly heterozygous apomictic complexes [49]. However, if it is true that the occurrence of apomixis in diploid *J. curcas* accessions is unlikely, it is also true that adventitious embryony cannot be ruled out in this species. According to such asexual mode of reproduction, embryos are initiated directly from individual cells in ovule tissues that are external to the sexually derived embryo sac. In fact, adventitious embryony has been described also as sporophytic apomixis [50]. Remarkably, genera with adventitious embryony are commonly diploid [49]. Conclusive results may only be gathered by means of cyto-histological analyses using sectioning or stain-clearing techniques as tools to shed light on sporogenesis, gametogenesis and embryogenesis pathways and its distinctive features in *J. curcas*.

Strong genomic uniformity among commercial varieties, as revealed by RAPD and Inter-SSR analyses, and a high degree of homozygosity as detected by SSR markers, are consistent with prevalent autogamy, the only reproductive pathway leading to a high degree of homozygosity [51]. As an alternative, an equally likely explanation for the low genetic variation observed over all accessions would be the repeated occurrence of bottlenecks that may have arisen through time and that may have caused the loss of any polymorphism. In accordance with this explanation, the low proportion of polymorphisms observed with all types of DNA markers could also be explained by inbreeding. In fact, marker data proving narrow genetic differentiation among varieties and low genic variation within varieties are in agreement not only with the expression of phenotypic variability for several plant traits

because of environmental factors, but also with the evidence of homozygosity for most of the genomic loci over all plants. The very low genetic variation in commercial varieties of *J. curcas* could also be explained by clonality owing to the use of vegetative cuttings as main strategy of propagation [52-54].

The main population genetic parameters support monogenotypic varieties attributable to clonality and prevalent homozygosity in landraces is likely due to a dominant inbreeding strategy of seed formation for *J. curcas*. As a matter of fact, with reference to commercial varieties, the Nei's estimate of genetic diversity was as low as  $H = 0.2597$  and the mean genetic similarity estimate as high as  $SM = 96.5\%$ . Additional co-dominant marker loci (e.g., SSR and SNP markers) and locally dominant ecotypes (e.g., Central America landraces) need to be tested in order to get conclusive results on the organization and distribution of gene polymorphisms at the population level.

## 5. Conclusions

The main finding is that seeds commercialized worldwide seem to be unrepresentative of the original gene pool, showing very low genetic variation either within or between varieties. Owing to the high degree of homogeneity found for most of the microsatellite loci across all accessions, commercial varieties of *J. curcas* consist of few genetically identical or closely related clones. Moreover, the clustering of accession groups was not correlated with their geographical area of cultivation and collection. The incontrovertible evidence that only eight dominant highly correlated multi-locus genotypes, tightly grouped into two main clusters, were found out of the 16 commercial varieties (44 accessions) suggests that a given variety can be distributed under different names. This result is explainable taking into account that certain *J. curcas* genotypes/clones introduced in African and Asian countries during last decades with a given name may have changed their phenotype under different selective environmental and anthropological pressures that most likely led to a variation of their morphological traits and then led final users to rename them. Additionally, misleading local nomenclature and incorrect germplasm recordkeeping may have contributed over the years to an erroneous or confused classification of varieties. The fact that most Indian accessions revealed heterozygosity for different microsatellite alleles at some of the loci investigated may reflect the development of commercial hybrids by controlled crosses between different clonal genotypes.

It is known that *J. curcas* has never been bred for productivity and/or quality traits and all commercial varieties derive by phenotype-based selections from local germplasm and are propagated by vegetative cuttings. As a consequence, narrow genetic base of *J. curcas* populations outside the centre of origin could probably be due to limited introductions without further augmentation of the crop, predominant vegetative propagation or occurrence of asexual reproduction (apomixis). Nevertheless, gametophytic apomixis does not seem to take place in *J. curcas* because all seeds revealed diploid embryo combined with triploid endosperm. Adventitious embryony cannot be discarded in this species because genera showing this reproductive strategy are commonly diploid. Conclusive results may only be gathered by means of cyto-histological analyses using sectioning or stain-clearing techniques as tools to shed light on sporogenesis, gametogenesis and embryogenesis pathways and its distinctive features in *J. curcas*.

As future perspectives, the genotyping of single plants and haplotyping of candidate genes using a larger *J. curcas* worldwide collection of local ecotypes will be attempted to find out the most

representative and discriminate polymorphisms (SNPs) eventually related with fatty acid biosynthesis and lipid metabolism genes. A major goal will be that of discovering association between specific alleles and seed oil contents and/or patterns. Finally, taking into account the whole set of experimental data, the genotypes showing the best quali-quantitative lipid profile will be propagated and used as materials for breeding new varieties suitable for biomass production by means of *in vitro* somatic embryogenesis.

## References

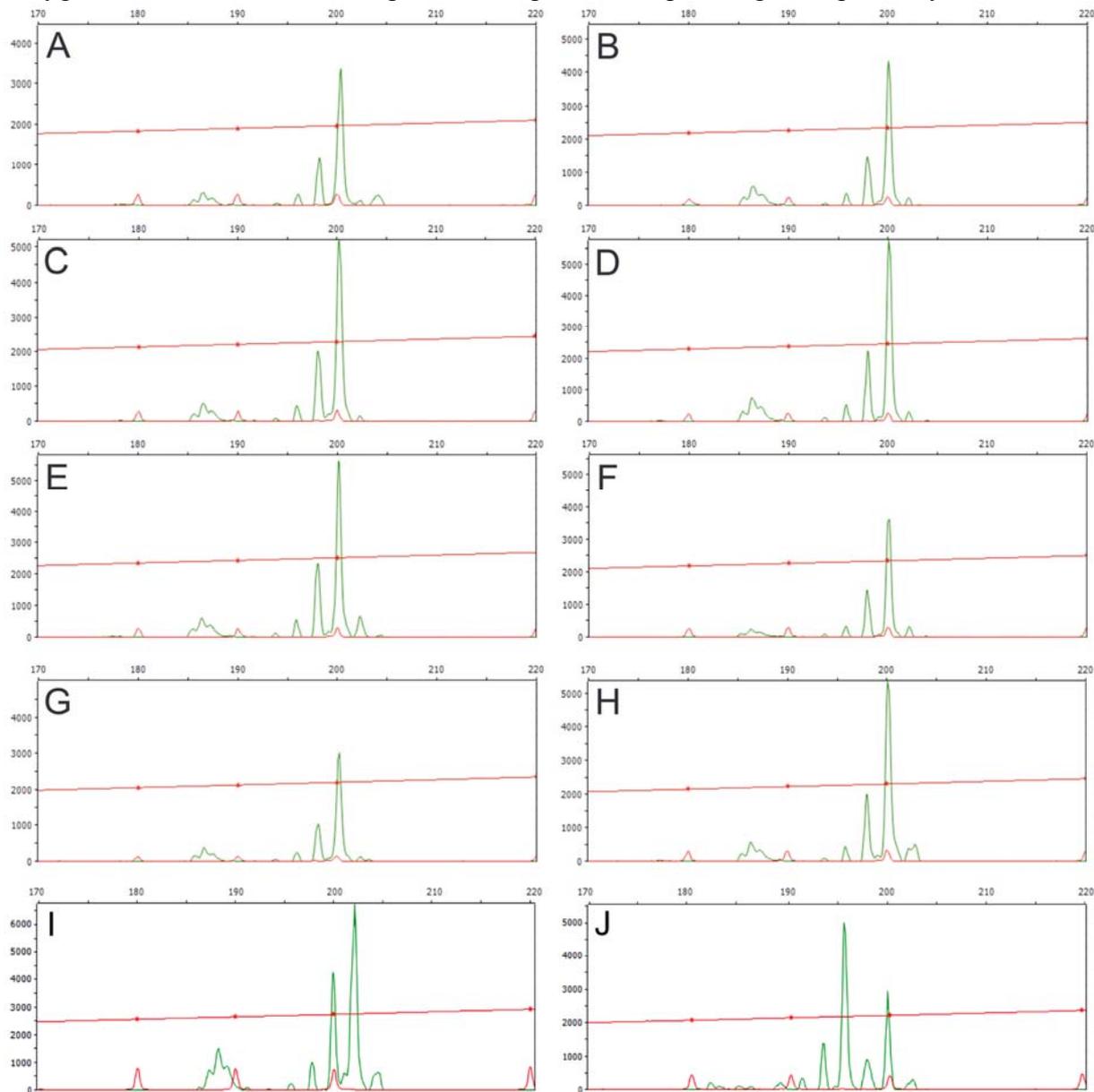
1. Zieler, H.; Richardson, T.; Schwartz, A.; Herrgard, M.; Lomelin, D.; Mathur, E.; Cheah, S.C.; Tee, T.S.; Lee, W.W.; Chua, K.L.; Kwan, B.H.; Miller, J.; Koren, S.; Chan, A.P.; Maiti, R.; Amedeo, P.; Ferriera, S.; Rogers, Y.H.; Sutton, G.; Town, C.D.; Venter, J.C. Whole-genome shotgun sequencing of the oil palm and *Jatropha* genomes. In Proceedings of the Plant and Animal Genomes XVIII Conference. Sherago International, Inc.: San Diego, CA, USA, January 2010; W. 360.
2. Nature Editorial. Kill king corn. *Nature* **2007**, *449*, 637.
3. Foidl, N.; Foidl, G.; Sanchez, M.; Mittelbach, M.; Heckel S. *Jatropha curcas* L. as a source for the production of biofuel in Nicaragua. *Bioresource Technol.* **1996**, *58*, 77-82.
4. Gubitz, G.M.; Mittelbach, M.; Trabi, M. Exploitation of the tropical oil seed plant *Jatropha curcas* L. *Bioresource Technol.* **1999**, *67*, 73-82.
5. Francis, G.; Edinger, R.; Becker, K. 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.* **2005**, *29*, 12-24.
6. Berchmans, H.J.; Hirata, S. Biodiesel production from crude *Jatropha curcas* L. seed oil with a high content of free fatty acids. *Bioresource Technol.* **2008**, *99*, 1716-1721.
7. Kumar A.; Sharma S. An evaluation of multipurpose oil seed crop for industrial uses (*Jatropha curcas*): A review. *Ind. Crops Prod.* **2008**, *28*, 1-10.
8. King, R.D.; Rowland, J.; Oliver, S.G.; Young, M.; Aubrey, W.; Byrne, E.; Liakata, M.; Markham, M.; Pir, P.; Soldatova, L.N.; Sparkes, A.; Whelan, K.E.; Clare, A. The automation of science. *Science* **2009**, *324*, 85-89.
9. Basha, S.D.; Francis, G.; Makkar, H.P.S.; Becker, K.; Sujatha, M. A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries. *Plant Sci.* **2009**, *176*, 812-823.
10. Ambrosi, D.G.; Galla, G.; Purelli, M.; Barbi, T.; Fabbri, A.; Barcaccia, G. DNA markers and FCSS analyses shed light on the genetic diversity and reproductive biology of *Jatropha curcas* L. In Proceedings of the Plant and Animal Genomes XVIII Conference. Sherago International, Inc., San Diego, CA, USA, January 2010; p. 129.
11. Spillane, C.; Curtis, M.D.; Grossniklaus, U. Apomixis technology development. Virgin births in farmers fields. *Nat. Biotechnol.* **2004**, *22*, 687-691.
12. Bhattacharya, A.; Datta, K.; Datta, S.K. Floral biology, floral resource constraints and pollination limitation in *Jatropha curcas* L. *Pakistan J. Biological Sci.* **2005**, *8*, 456-460.

13. Luo, C-W.; Li, K.; Chen Y.; Sun Y.Y. Floral display and breeding system of *Jatropha curcas* L. *Forestry Stud. China* **2007**, *9*, 114-119.
14. Popluechai, S.; Breviario, D.; Mulpuri, S.; Makkar, H.P.S.; Raorane, M.; Reddy, A.R.; Palchetti, E.; Gatehouse, A.M.R.; Syers, J.K.; O'Donnell, A.G.; Kohli, A. Narrow genetic and apparent phenetic diversity in *Jatropha curcas*: Initial success with generating low phorbol ester interspecific hybrids. *Nature Proceedings* **2009**, (available at hdl: 10101/npre.2009.2782.1).
15. Pamidimarri, D.V.N.S.; Singh, S.; Mastan, S.G.; Patel, J.; Reddy, M.P. 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.* **2009**, *36*, 1357-1364.
16. Basha, S.D.; Sujatha, M. Inter and intra-population variability of *Jatropha curcas* L. characterized by RAPD and ISSR markers and development of population-specific SCAR markers. *Euphytica* **2007**, *156*, 375-386.
17. Yu, C.; Wu, G.; Peng, J. ISSR-based genetic diversity of *Jatropha curcas* germplasm in China. In Proceedings of the Plant and Animal Genomes XVIII Conference. Sherago International, Inc.: San Diego, CA, USA, January 2010; P. 212.
18. Pamidimarri, D.V.N.S.; Nirali, P.; Reddy, M.P.; Radhakrishnan, T. Comparative study of interspecific divergence and phylogenetic analysis of genus *Jatropha* by RAPD and AFLP. *Mol. Bio. Rep.* **2008**, *36*, 901-907.
19. Doebley J.F.; Gaut B.S.; Smith, B.D. The molecular genetics of crop domestication. *Cell* **2006**, *127*, 1309-1321.
20. Kaushik, N.; Kumar, K.; Kumar, S.; Kaushik, N.; Roy, S. Genetic variability and divergence studies in seed traits and oil content of *Jatropha* (*Jatropha curcas* L.) accessions. *Biomass and Bioenergy* **2007**, *31*, 497-502.
21. Popluechai, S.; Raorane, M.; Syers, K.J.; O'Donnell, A.G.; Kohli, A. Research needs to make *Jatropha* a viable alternate oilseed crop. In *Proceedings of the International workshop on Feasibility of non-edible oil seed crops for biofuel production*, Mae Fah Luang University; Syers, K.J., Wood, D., Eds.; MFLU publication: Chiang Rai, Thailand, May 2007, pp. 23-38.
22. Gressel, J. Transgenics are imperative for biofuel crops. *Plant Sci.* **2008**, *174*, 246-263.
23. Matzk, F.; Meister, A.; Schubert, I. An efficient screen for reproductive pathways using mature seeds of monocots and dicots. *Plant J.* **2000**, *21*, 97-108.
24. Dolezel, J.; Binarova, P.; Lucretti, S. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* **1989**, *31*, 113-120.
25. Mendonça, M.A.C.; Carvalho, C.R.; Clarindo, R.W. DNA content differences between male and female chicken (*Gallus gallus domesticus*) nuclei and Z and W chromosomes resolved by image. *J. Histochem. Cytochem.* **2009**, *58*, 229-235.
26. Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed.; NY Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 1989; pp. 1659.
27. Barcaccia, G.; Mazzucato, A.; Belardinelli, A.; Pezzotti, M.; Lucretti, S.; Falcinelli, M. Inheritance of parental genomes in progenies of *Poa pratensis* L. from sexual and apomictic genotypes as assessed by RAPD markers and flow cytometry. *Theor. Appl. Genet.* **1997**, *95*, 516-524.

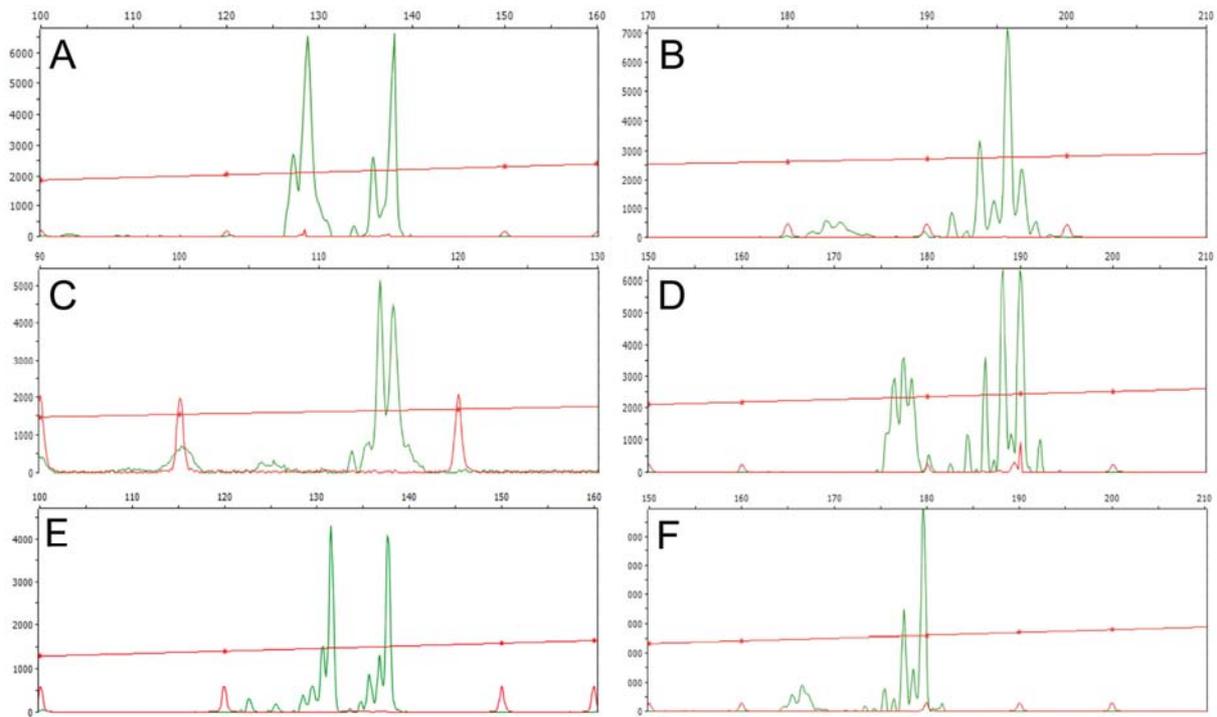
28. Barcaccia, G.; Lucchin, M.; Parrini, P. Characterization of a flint maize (*Zea mays* var. *indurata*) Italian landrace: II. Genetic diversity and relatedness assessed by PCR-based molecular markers. *Genet. Res. Crop Evol.* **2003**, *50*, 253-271.
29. Barcaccia, G.; Meneghetti, S.; Albertini, E.; Triest, L.; Lucchin, M. Linkage mapping in tetraploid willows: segregation of molecular markers and estimation of linkage phases support an allotetraploid structure for *Salix alba* x *Salix fragilis* interspecific hybrids. *Heredity* **2003**, *90*, 169-180.
30. Hayden, M.J.; Nguyen T.M.; Watrman, A.; McMichael, G.L.; Chalmers, K.J. Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Mol. Breeding* **2008**, *21*, 271-281.
31. Zubietta, C.G.; Ghiselli, L.; Benedettelli, S.; Palchetti, E. Development of novel SSR markers from genomic microsatellite library in *Jatropha curcas* L. In Proceedings of the 53rd Italian Society of Agricultural Genetics Annual Congress. SIGA publisher: Torino, Italy, September, 2009; p. 7, 60.
32. Kimura, M.; Crow, J.F. The number of alleles that can be maintained in a finite population. *Genetics* **1964**, *49*, 725-738.
33. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 3321-3323.
34. Lewontin, R.C. *The Genetic Basis of Evolutionary Change*; Columbia University Press: New York, NY, USA, 1974.
35. McDermott, J.M., McDonald, B.A. Gene flow in plant pathosystems. *Annu. Rev. Phytopathol.* **1993**, *31*, 353-373.
36. Wright, S. The interpretation of population structure by F-statistics with special regards to systems of matings. *Evolution* **1965**, *19*, 395-420.
37. Wright, S. *Evolution and the Genetics of Populations. Variability Within and Among Natural Populations*; University of Chicago Press: Chicago, IL, USA, 1978; vol. 4.
38. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **1995**, *139*, 457-462.
39. Goodman, S.J. Rst Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Mol. Ecol.* **1997**, *6*, 881-885.
40. Yeh, F.C.; Yang, R.C.; Boyle, T. *POPGENE version 1.21*; CIFOR and University of Alberta: Edmonton, Canada. 1997.
41. Sneath, P.H.A.; Sokal, R.R. *Numerical Taxonomy*; Freeman W.H. Co: San Francisco, CA, USA, 1973.
42. Gower, J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **1996**, *53*, 325-338.
43. Rohlf, F.J. *NTSYS-pc: Numerical Taxonomy System (version 2.1)*; Exeter Publishing, Ltd.: Setauket, NY, USA, 2002.
44. Pritchard, J.K.; Stephens, P.; Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945-959.

45. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **2003**, *164*, 1567-1587.
46. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **2005**, *14*, 2611-2620.
47. Carvalho, C.R.; Clarindo, W.R.; Prac, M.M.; Araujo, F.S.; Carels, N. Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. *Plant Sci.* **2008**, *174*, 613-617.
48. Galbraith, D.W.; Harkins, K.R.; Maddox, J.M.; Ayres, N.M.; Sharma, D.P.; Firoozabady, E. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **1983**, *220*, 1049-1051.
49. Loureiro, J.; Rodriguez, E.; Dolezel, J.; Santos C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Annals Botany* **2006**, *98*, 679-689.
50. Savidan, Y.H. Apomixis: Genetics and breeding. *Plant Breed.* **2000**, *18*, 10-86.
51. Barcaccia, G.; Baumlein, H.; Sharbel, T.F. Apomixis in St. John's wort: An overview and glimpse towards the future. In *Apomixis: Evolution, Mechanisms and Perspectives*; Hörandle, E., Grossniklaus, U., Van Dijk, P., Sharbel, T.F., Eds.; Intl. Association of Plant Taxonomy - Koeltz Scientific Books: Vienna, Austria, 2007; Chapter XIV, pp. 259-280.
52. Asker, S.E.; Jerling, L. *Apomixis in plants*; CRC Press: London, UK, 1992.
53. Nogler, G.A. Gametophytic apomixis. In *Embryology of Angiosperms*; Johri, B.M., Ed.; Springer-Verlag: Berlin, Germany, 1984; pp. 475-518.
54. Allard, R.W. History of plant population genetics. *Annu. Rev. Genet.* **1999**, *33*, 1-27.
55. Heller, J. *Physic Nut. Jatropha curcas* L. *Promoting the Conservation and Use of Underutilized and Neglected Crops*. Institute of Plant Genetics and Crop Plant Research (IPK), International Plant Genetic Resources Institute (IPGRI); Gatersleben, Germany, Rome, Italy, 1996; pp. 65.
56. Jongshaap, R.E.E.; Corré, W.J.; Bindraban, P.S.; Brandenburg, W.A. *Claims and Facts on Jatropha curcas* L. *Global Jatropha curcas Evaluation, Breeding and Propagation Programme*; Report 158; Plant Research International B.V., Wageningen Stichting Het Groene Woudt: Laren, The Netherlands, 2007.
57. Montes Osorio, L.R.; Van Loo, E.N.R.; Jongshaap, R.E.E.; Visser, R.G.H.; Azurdia, C. Global *Jatropha curcas* Genetic Diversity Study and its Application in Breeding Programs, In *JatrophaWorld 2008: An OUTSTANDING Event – Two Thumbs Up*, Miami, FL, USA, June 2008.

**Appendix 1.** DNA fragment profiles showing microsatellite loci assayed with primer JcSSR7 using eight different *J. curcas* accessions as genomic DNA templates. **A**, SLK2; **B**, BRA1; **C**, BER1; **D**, TOG3; **E**, LOL1; **F**, SEL2; **G**, ENG3; **H**, PET1; **I**, CLO4; and **J**, HDE4. All varieties proved to be homozygous for a 200 bp long allele, whereas the Mexican accessions revealed both homozygous and heterozygous loci for alleles of 202 bp, and 196 bp and 200 bp in length, respectively.



**Appendix 2.** DNA fragment profiles showing two different alleles in each of the heterozygous loci investigated. **A**, JcSSR2; **B**, JcSSR4; **C**, JcSSR9; **D**, JcSSR3; **E**, JcSSR2, and **F**, JcSSR10.



© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).