

Study of genetic diversity and differentiation of argan tree population (*Argania spinosa* L.) using AFLP markers

Ouafae Pakhrou¹, Leila Medraoui¹, Chaimaa Yatrib¹, Mohammed Alami¹, Saad Ibn Souda-kouraichi², Abdelhamid El mousadik³, Abderrahim Ferradous⁴, Fouad Msanda³, Cherkaoui El modafar⁵, Abdelkarim Filali-maltouf^{1*}, Bouchra Belkadi¹

¹Laboratory of Microbiology and Molecular Biology, Faculty of Science, Mohammed Vth 8 University, Rabat, Morocco

²Laboratory of Microbial Biotechnology, Faculty of Sciences and Technics, Sidi Mohammed Ben Abdellah University, Fez, Morocco

³Laboratory of Biotechnologies and Valuation of Natural Resources, Ibn-Zohr University, Faculty of Science, Agadir, Morocco

⁴High Commission for Water Forests and Desertification Control (HCEFLCD), Forestry Research Center (FRC), Marrakech, Morocco

⁵Laboratory of Biotechnology, for the Agro-ressources Valorization and the Protection, Faculty of Sciences and Technics, Cadi Ayyad University, Marrakech, Morocco

*Corresponding author: a.filalimaltouf@gmail.com

Abstract

The argan tree (*Argania spinosa* L. Skeels, Sapotaceae) is a genetic resource endemic in Morocco. Genetic diversity within and among 13 populations (130 genotypes) of argan tree was studied using AFLP markers. Having checked twenty combinations of labeled primers for regular genomes (500-6000 Mb) (EcoRI+3/ MseI+3 selective bases) and for small genome (50-500 Mb) (EcoRI+2/MseI+3 selective bases), we selected four combinations specific for regular genome able to produce a relatively high polymorphism and a low error rate (0.12 %). A total of 477 unambiguous peaks were amplified ranging from 70 to 500 bp. The average polymorphism information content (PICA_{AVG}) value ranged from 0.19 to 0.23. Marker index (MI) and resolving power (RP) varied from 21.23 to 28.82 and 27.63 to 44.92, respectively. Analysis of molecular variance (AMOVA) showed that 19 % of the genetic variation was partitioned among populations and 81 % of the genetic variation was within populations. This was confirmed by the coefficient of gene differentiation between populations ($G_{st}=0.22$), and gene flow was estimated to 1.709. The STRUCTURE analysis, principal coordinate analysis (PCoA) and Unweighthed Pair Group Method with Arithmetic Mean (UPGMA) revealed that populations of *A. spinosa* were clustered into three genetic groups. The present results can be explored in the design of in situ and ex situ conservation and management programs.

Keywords: AFLP; *Argania spinosa* L.; Genetic differentiation; Genetic diversity.

Abbreviations: AMOVA_analysis of molecular variance; G_{st} _coefficient of gene differentiation; MI_marker index. PCoA_principal coordinates analysis; PIC_polymorphism information content; RP_resolving power; UPGMA_Unweighthed Pair Group Method with Arithmetic Mean.

Introduction

The argan tree is exclusively endemic to Southwestern Morocco, which is perfectly adapted to the soils and climate of the region. It is the only species of the genus *Argania* representing the family Sapotaceae in the subtropical areas. The tree survival in this region is due to its deeper root system and ability to drop leaves in severe drought conditions (Msanda et al., 2007).

This tree plays an undeniable role against erosion and desertification and furnishes, via its fruit, an oil high in unsaturated fatty acids (Chimi et al., 1994) in great demand worldwide for medicinal, cosmetic and nutritional virtues (Lybbert et al., 2010). Moreover, it contributes to the income of rural families. In fact, it has been considered as a national

resource with a high ecological and socio-economic impact (Mhirit et al., 1998).

Unfortunately, the area and density of the argan tree forest have suffered from many serious depletion, caused by a combination of loss of trees and lack of regrowth (le Polain de Waroux and Lambin, 2012) due to overgrazing, the increasing aridity and the excessive fuel wood extraction (Nouaim, 2005; Naggat and Mhirit, 2006). The degradation of argan forests at an alarming rate causes potential loss of the existing genetic diversity, which needs quite urgent action to better manage and safeguard the remaining. Thus, the conservation of genetic variation is a necessary element to the sustainability and the adaptation of the species to their changing environments (Futuyma, 1986; Namkoong et al.,

2002). In order to conserve genetic resources, the analysis of the structure and the understanding of distribution patterns of genetic diversity are primordial for all programs (Hamrick and Godt, 1996). To date, most of *A. spinosa* L. studies have been mainly focused on its fruit and oil (Alaoui et al., 1998; Bani-Aameur et al., 1999; Charrouf and Guillaume, 1999; Charrouf et al., 2002; Sparg et al., 2004; Harhar et al., 2014). However, there are few investigations that show basic information regarding the genetic variation within/among populations *A. spinosa* L. (El Mousadik and Petit, 1996; Majourhat et al., 2008), and AFLP markers have, to the best of our knowledge, never been used to analyze the genetic diversity of *A. spinosa* L.

The AFLP method is a dominant molecular marker considered a powerful tool for genetic analysis (Vos et al., 1995; Bensch and Akesson, 2005; Ni et al., 2006) that allows for analysis of a large number of marker loci in the genome without needing prior knowledge of DNA sequences. It generates a high genetic polymorphism and has good reproducibility (Mueller and Wolfenbarger, 1999; Lamote et al., 2002). This marker system has found wide applications in population genetic studies of various species e.g. in *Picea chihuahuana*/Pinaceae (Simental-Rodríguez et al., 2014), *Paulownia fortune*/Paulowniaceae (Li et al., 2014), *Rhododendron protistum*/Ericaceae (Wu et al., 2014).

In this work, AFLP markers were used to evaluate the genetic diversity and estimate the genetic differentiation among and within 13 populations of *A. spinosa* L. collected across Essaouira province in south-west of Morocco. The obtained information will contribute to the conservation and maintenance of such a genetic resource.

Results

Marker polymorphism, AFLP features, and fragment distribution

To evaluate and characterize the 130 argan genotypes, a total of twenty AFLP primer combinations were screened. All primer combinations specific to small genome generated a negative or monomorphic AFLP profiles, for both tomato and *A. spinosa* L. DNA, and positive profiles for Arabidopsis genome. While ten specific regular genome combinations successfully produced scorable, clear high-intensity, reproducible and a moderate number of bands (Fig. 1). Four out of ten primer combinations, specific to regular genome, producing well-separated peaks and relatively high polymorphism (Fig. 2), were selected for the present study. The error rate of these four primer combinations was estimated to 0.12 %. These combinations allowed the detection of 477 AFLP loci in total. The total number of fragments for each primer combination varied from 109 (E-ACC/M-CTC) to 126 (E-ACC/M-CAT), with a mean of 119.25. Likewise, the percentage of polymorphism was a 100% for all the primer combinations.

The powerfulness and discriminatory power of the informative AFLP profile mainly depends on three parameters namely polymorphic information content (PIC), marker index (MI) and resolving power (RP). The calculated PIC values ranged from 0.19 to 0.23 with an average of 0.21 per primer combinations (Table 1). The primer combination E-ACC/M-CAT was the most informative according to the PIC value and the lowest PIC value (0.19) was noticed for E-ACC/M-CTC. The marker index (MI) varied from 21.23 (E-ACC/M-CTC) and 28.82 (E-ACC/M-CAT) with an average

of 25.61. Resolving power (RP) ranged from 27.63 (E-ACC/M-CTC) to 44.92 (E-ACC/M-CAG) with an average of 35.81. A strong positive correlation was noticed between RP and PIC ($r=0.74$; $p>0.05$), between MI and PIC value ($r=0.95$, $p >0.05$), and between RP and MI ($r=0.57$; $p>0.005$). Based on the distribution frequency, the AFLP generated fragments were classified into unique fragments (NUF), rare fragments (NRF) similar fragments (NSIF) and shared fragments (NSF). A total of 30 (6%) unique fragments were identified varying from 5 to 12 with an average of 7.5 per PC. The highest number of unique fragments was produced by E-ACC/M-CTC. The AFLP fragments that occur in less than 10% of genotypes used in the study were considered as rare fragments. A total of 266 (56%) rare fragments were identified and varied from 60 (E-ACC/M-CTC) to 71 (E-ACC/M-CAG) with an average of 66.5 rare fragments per PC. Similar fragments are those AFLP fragments scored in more than 70% genotypes. A total of 34 (7%) similar fragments were recorded varying from 6 (E-ACC/M-CAG) to 13 (E-ACC/M-CTA) with an average of 16.2 per PC. The AFLP fragments for a particular locus found between 10% and 70% of genotypes were considered as 'shared fragments'. Moreover, a total of 147 (31%) shared fragments were recorded varying from 30 (E-ACC/M-CTC) to 44 (E-ACC/M-CAT) with an average of 36.75.

Genetic diversity and population structure analysis

The analysis of molecular variance (AMOVA) showed that 19 % of the total genetic variation exists among the populations and 81% within population (Table 2). The *Gst* was 0.22, indicating that 22% of total genetic variability was among populations and 78% was within populations. The gene flow (*Nm*) calculation based on the *Gst* value between populations was found to be 1.709 (Table 2). The genotypes used in the present investigation showed a moderate level of variation as indicated by the ranges: *Na* (0.998- 0.612), *Ne* (1.192 - 1.123), *H* (0.120 - 0.075) and *I* (0.195 - 0.118) (Table 3).

The three coordinates, which were obtained according to the PCoA (Fig. 3), explained 48.13 % of total variation. The first axis displayed 28% of the variance, while the second and the third ones explained 14.47% and 5.67% of the variance respectively. This analysis presents three clear groups of individuals: group 1 is formed by the population of Mramer, group 2 by Ait Issi, Neknafa, Tamesart, Timzgida Oufetass and Lakhsas and group 3 by Imouzzar. The remaining populations are dispersed mainly in the two latter groups. Jaccard's similarity coefficients among different populations varied from 0.14 to 0.69 with an average of 0.48 (Fig. 4) suggesting the presence of moderate genetic diversity. UPGMA dendrogram presents the genetic relationships among different populations which showed that all 130 genotypes of *A. spinosa* L. are grouped into three major clusters.

A bayesian analysis was performed based on the molecular data in order to confirm the results obtained with the UPGMA clustering and PCoA described above that provided further evidence of hierarchical populations structure and identify the number of the most appropriate populations.

The second order statistics (ΔK) developed by Evanno et al. (2005) for STRUCTURE, determined the optimal value for *K*=3 (Fig. 5). These results confirm the three clustering that we observed in UPGMA dendrogram and PCoA.

Table 1. Degree of polymorphism and information content for four AFLP primer combinations used.

Primer combinations (PC)	TF ^a	PP ^b (%)	NUF ^c	NRF ^d	NSF ^e	NSIF ^f	PIC ^g	RP ^h	MI ⁱ
E-ACC/M-CTC	109	100	12	60	30	7	0.19	27.63	21.23
E-ACC/M-CTA	124	100	6	66	39	13	0.21	34.14	26.51
E-ACC/M-CAT	126	100	5	69	44	8	0.23	36.55	28.82
E-ACC/M-CAG	118	100	7	71	34	6	0.22	44.92	25.88
Average	119		7.5	66.5	36.75	16.2	0.21	35.81	25.61

^a Total number of fragments; ^b Percent polymorphism; ^c Number of unique fragments; ^d Number of rare fragments; ^e Number of shared fragments; ^f Number of similar fragments; ^g Polymorphic information content; ^h Resolving power; ⁱ Marker Index.

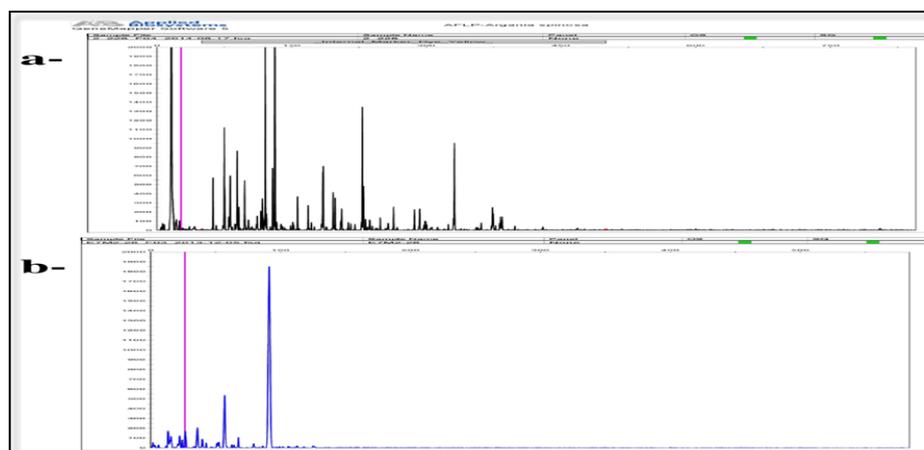


Fig 1. Representative electropherograms (snapshot from GeneMapper) showing AFLP profiles of *A. spinosa* L., with primer specific to small genome generated only monomorphic bands (a-) and primer specific to regular genome successfully produced scorable, clear high-intensity, and a moderate number of bands (b-).

The first cluster (red color) consisted mainly of Mramer population, the other groups (green color and blue color) reveal the presence of other mixed individuals from different populations and show similar distribution of PCoA and UPGMA.

Discussion

The quality of AFLP fingerprints depends on the size of the genome and the C-value (FAY et al., 2005). The missing data on the argan tree genome size forces us to test two protocols recommended by Applied Biosystems, 1996, namely the specific protocol for small genomes (50-500 Mb) (2+3 selective bases) and the specific protocol for regular genomes (500-6000 Mb) (3+3 selective bases). The specific primers for small genomes did not generate any polymorphism except monomorphic bands. However, the specific primers for regular genomes are the most compatible with the argan tree genome, generating enough fragments to compare samples. The revealed polymorphism is similar to the generated polymorphism in the tomato (positive control) and different of the Arabidopsis profiles. It is difficult to estimate the content of the argan tree's nuclear DNA based on the detected polymorphism but the poor quality of peaks obtained by the specific protocol for small genomes suggest that the genome size of *A. spinosa* L. exceeds 500 Mb.

Furthermore, being aware of possible errors in the ratings of AFLP bands, we re-analyzed one individual from each population. The bands produced by the four combinations of primers selected to study the genetic diversity of *A. spinosa* L. showed repetitive patterns and detected low error rate (0.12%). These results reinforce the quality of the analysis and the revealed polymorphism. The estimated error rate in our analysis is lower than in other studies. For instance, a higher error rate was observed by Bonin et al. (2004). In

particular, they found that the error rate of 2.6% was associated with AFLP analysis of genomic DNA extracted from *Betula nana*.

The fingerprinting of the AFLP profile of PCR products of 130 genotypes, for primer combinations, revealed a total number of 477 unambiguous polymorphic amplified DNA fragments. This confirms that with a small number of primer combinations, AFLP can be used to detect substantial numbers of polymorphic loci (Balasaravanan et al., 2003). In *Acacia* the use of four AFLP primer combinations allowed the identification of 228 polymorphic loci (Pometti et al., 2015) and 250 polymorphic reproducible loci were revealed in the *Schinus molle* of only two AFLP primer combinations (Lemos et al., 2015); making the AFLP a powerful tool for genetic variability and population structure studies (Mueller and Wolfenbarger, 1999; Bensch and Akesson, 2005). The analysis of the frequency and the distribution of AFLP fragments among all individuals have shown that the combination of primer E-ACC/M-CTC generated the maximum number of unique fragments (12 unique fragments) and E-ACC/M-CAG produced 7 unique fragments. These primer combinations can be used to develop Sequence Tagged Sites (STS) markers in order to identify particular genotypes (Fernández et al., 2002). It is interesting to note that the largest number of unique fragments was observed in the Mramer population. The presence of unique alleles can also reveal important information about the differentiation of the genetic structure of a population. The new alleles in a population occur as a result of mutations (Mariette et al., 2002) showing the adaptive capacity of the species in its environment. Mramer population is a geographically remote from the rest of the studied populations. Therefore, the number of unique fragments in this population generates genetic differentiation between Mramer population and other populations as observed (Fig. 1).

Table 2. Analysis of molecular variance (AMOVA) of 13 populations of *A. spinosa* L.

Source	Degrees of freedom	Sum of squares	Mean of squares	Variance component	% Variance component	G_{st} ^a	N_m ^b
Among population	12	1343.631	111.969	7.782	19%	0.22	1.709
Within population	117	3995.400	34.149	34.149	81%		
Total	129	5339.031		41.931	100%		

^a Coefficient of genetic differentiation; ^b Estimate of gene flow.

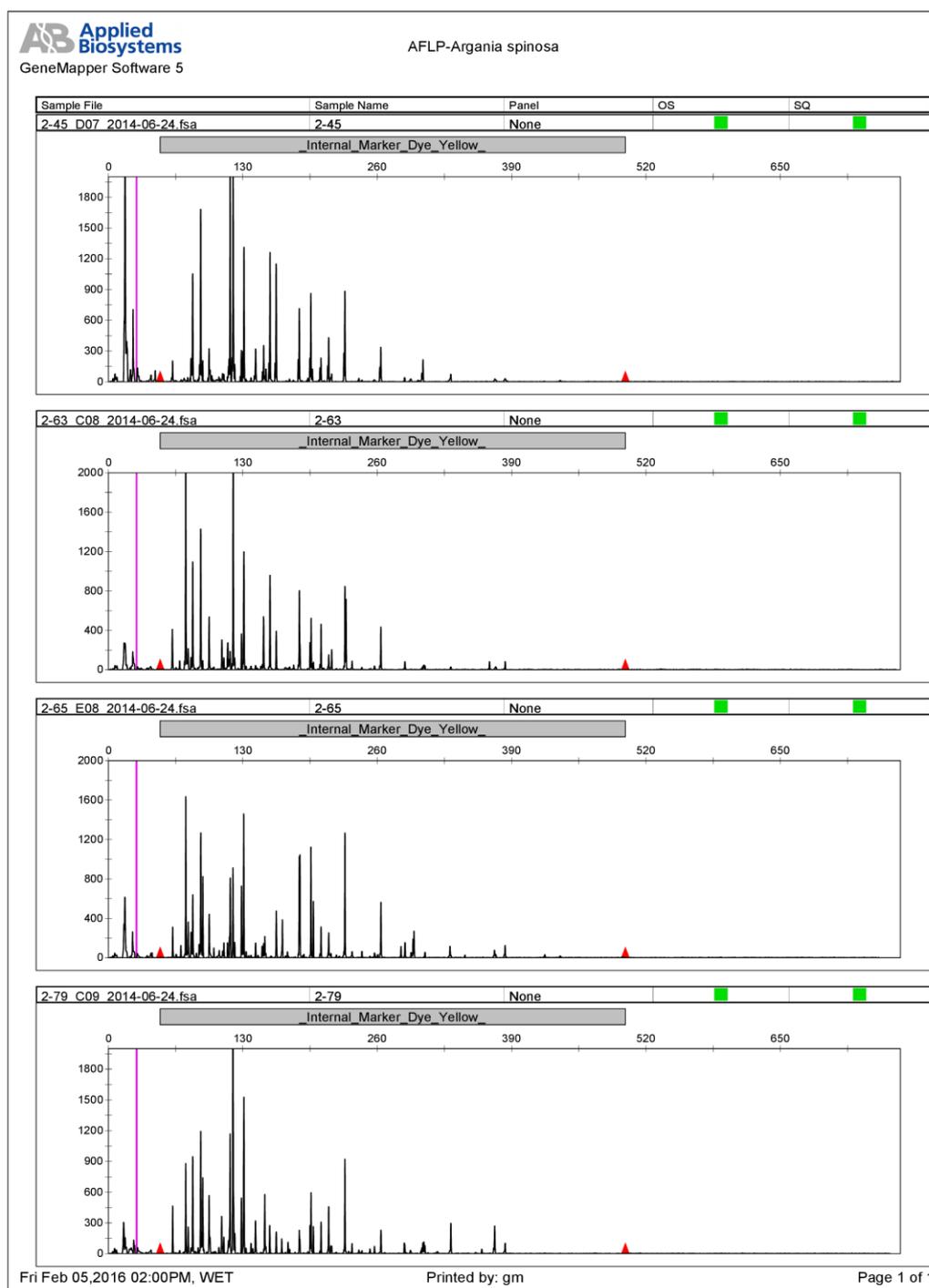


Fig 2. Representative electropherograms (snapshot from GeneMapper) showing AFLP profiles of four individuals of *A. spinosa* L. (Mramer population), with primer combination E-ACC/M-CTA (specific to regular genome) producing well-separated peaks and relatively high polymorphism.

Table 3. Genetic variation and polymorphic features estimated using AFLP markers among 13 populations of the *A. spinosa* L.

Populations	Na ^a	Ne ^b	I ^c	H ^d	%P (ppl) ^e
Mramer	0.960	1.192	0.189	0.120	46.12%
Rbai	0.998	1.184	0.185	0.116	47.59%
Tamesart	0.897	1.154	0.164	0.101	42.35%
Neknafa	0.891	1.147	0.156	0.095	41.72%
Ait Issi	0.612	1.123	0.118	0.075	27.88%
Tmanar	0.956	1.172	0.178	0.110	46.12%
Timzgidah Oufetass	0.874	1.161	0.164	0.102	42.56%
Admine	0.985	1.187	0.195	0.121	48.85%
Mnizla	0.885	1.178	0.179	0.113	42.35%
Imouzzer	0.648	1.163	0.149	0.098	30.40%
Tafraout	0.912	1.178	0.182	0.114	45.07%
Ait Baha	0.853	1.159	0.166	0.103	42.35%
Lakhssass	0.627	1.137	0.131	0.084	30.82%
Total	0.854	1.164	0.166	0.104	41.09%
Mean	0.854	1.164	0.166	0.104	

^a Observed number of alleles; ^b Effective number of alleles; ^c Shannon's information index; ^d Nei's (1973) gene diversity; ^e Percentage of Polymorphic Loci.

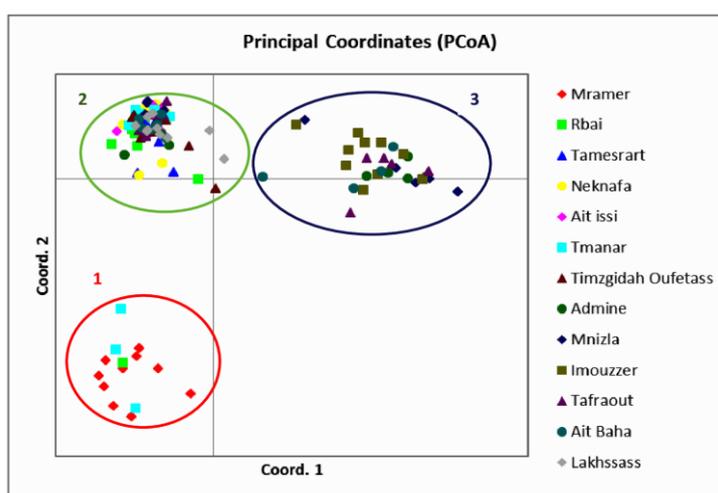


Fig 3. Principal coordinate analysis (PCoA) of 130 *A. spinosa* L. individuals from 13 populations, based on AFLP analyses by GenAlEx 6.5. Dispersion of 130 individuals shows three groups, group 1 (red) population of Mramer, group 2 (green) formed predominately by Ait Issi, Neknafa, Tamesart, Timzgidah Oufetass and Lakhssas; group 3 (bleu) containing essentially population of Imouzzer (the remaining populations are dispersed mainly in the two latter groups). Percentage of total variance was 48.13 %. (axis 1=28%; axis 2=14.47%; axis 3=5.67%).

The number of rare fragments produced by the four primer combinations is quite high compared to other plants (Tatikonda et al., 2009; Gupta et al., 2013). This means that the argan tree genome is rich in specific alleles to particular genotypes (Agrama and Tuinstra, 2003). Similar and shared fragments can be used in studies of evolution. They give information about the sympleisiomorphiques alleles (ancestral alleles) and identify the homologous regions in other taxa (Abdalla et al., 2001).

PIC, RP, and MI parameters have been used in several studies of plants genetic diversity to assess the discriminative power of AFLP markers (Baraket et al., 2009; Pecina-Quintero et al., 2013). The PIC values that we found are considered quite informative. They have a good discriminatory power and can be used to characterize the genetic material of the argan tree (Roldán-Ruiz et al., 2000). The other two parameters MI and RP have very high values indicating the discriminative potential of primer combinations used and show a positive correlation with PIC ($r=0.95$, $p>0.05$) and ($r=0.74$; $p>0.05$), respectively. The positive correlations observed between PIC/MI, MI/Rp and the large number of polymorphic bands generated (477) by these primers suggest that AFLPs are a highly

discriminative and powerful markers. Moreover it provides important information on the diversity analysis of *A. spinosa* L. (Prevost and Wilkinson, 1999; Fernández et al., 2002).

These parameters can also be explored to select combinations of informative primers. They might exhibit the highest values [E-ACC/M-CAT and E-ACC/M-CAG], which may be the most appropriate to study genetic mapping and phylogenetic analysis (Anderson et al., 1993; Powell et al., 1996). The primer combination [E-ACC/M-CTC] has the lowest PIC value. This is explained by the number of unique generated fragments.

According to Hamrick and Godt (1989), the coefficient of differentiation of plant species depends on the method of propagation, the geographical area and the size of the population. Thus, the widely distributed species have higher diversity within the population (Hamrick and Godt, 1989; Hogbin and Peakall, 1999). However, our results show that the argan tree with a restricted distribution has a higher differentiation than other forest trees: *Atlas Cypress* ($Gst=0.013$) (Bechir et al., 2004), *Quercus Robur* ($Gst=0.024$) (Scotti-Saintagne et al., 2004). These results are consistent with those found by El Mousadik and Petit (1996)

Table 4. Locations and sample size of 13 populations of *A. spinosa* L.

Population	Code	Sample size	Longitude	Latitude	Altitude (m)
Mramer	MR	10	9°10'	31°38'	396
Rbai	RB	10	9°28'	31°32'	269
Tamesart	TS	10	9°22'	31°21'	540
Neknafa	NK	10	9°33'	31°19'	242
Ait Issi	AI	10	9°22'	31°02'	980
Tmanar	TM	10	9°37'	31°00'	576
Timzgidah Oufetass	TO	10	9°48'	30°00'	231
Admine	AD	10	9°21'	30°19'	83
Mnizla	MN	10	9°05'	30°33'	256
Imouzzer	IZ	10	9°30'	30°39'	1100
Tafraout	TA	10	9°03'	29°42'	900
Ait Baha	AB	10	9°13'	30°06'	490
Lakhssass	LA	10	9°43'	29°24'	954

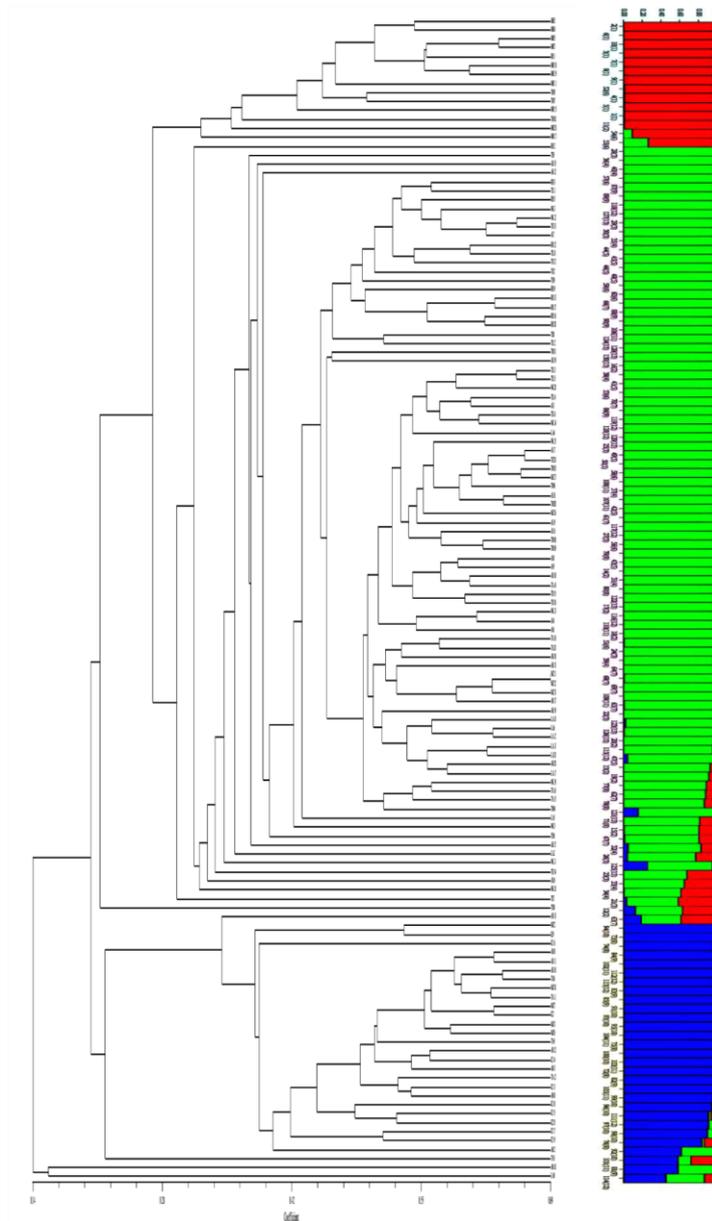
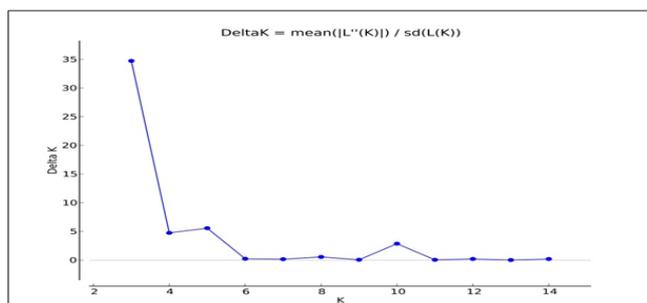


Fig 4. UPGMA-dendrogram based on Jaccards coefficient and hierarchical organization of genetic relatedness of *A. spinosa* L. analyzed by STRUCTURE program. A total of 130 trees showing assignation to three groups (Each color represents one group): Group 1 (red), group 2 (green), group 3 (bleu). Details regarding group 1, 2 and 3 are cited in the text and in the Fig. 2.

Table 5. List of primer combinations tested for regular and small genome.

Primer combinations -regular genome		Primer combinations - small genome	
E-ACC/M-CTC	E-AGG/M-CTC	E-AT/M-CTC	E-AC/M-CTC
E-ACC/M-CTA	E-AGG/M-CTA	E-AT/M-CTA	E-AC/M-CTA
E-ACC/M-CAT	E-AGG/M-CAT	E-AT/M-CAT	E-AC/M-CAT
E-ACC/M-CAG	E-AGG/M-CAG	E-AT/M-CAG	E-AC/M-CAG
E-ACC/M-CAA	E-AGG/M-CAA	E-AT/M-CAA	E-AC/M-CAA

**Fig 5.** The detection of the true number of clusters inferred by the STRUCTURE software; ΔK calculated as the mean of the second-order rate of change in likelihood of K divided by the standard deviation of the likelihood of K , $m|L'(K)|/sd[L(K)]$; where inferred clusters (K) ranged from 2 to 15. ΔK achieves its highest peak when $K=3$ (the most likely value of K); generated by the STRUCTURE HARVESTER, and based on the approach of Evanno et al. (2005).

($Gst=0.25$) using isoenzymes and by Ait Abd et al. (2015) ($Gst=0.22$) using ISSR markers. The estimation of the gene flow is also an important index to understand the genetic structure of populations. Gene flow between and within population was estimated to 1.709. This value (>1.0) is important to prevent genetic divergence in a population (Wright, 1951; Slatkin, 1987). Moreover, an exchange of genes, between studied populations, is expected for several reasons. Firstly, the studied sources are located in the same geographical area under some variable bioclimate (arid or semi-arid). Secondly, reproduction biology of the species and its pollen dispersal mechanism over long distances by wind and insects (Bani-Aameur et al., 1999). This result can be explained by the outcrossing mating system with anemophily and entomophily pollination of the argan tree. This creates a greater differentiation compared to wind-pollinated species (Hamrick and Godt, 1989; Aradhya and Phillips, 1993), than by disseminating seeds by animals despite the fragmented range of the argan tree. This result is confirmed by the different distribution structures studied (PCoA, UPGMA, STRUCTURE). In fact, many populations share their alleles with neighboring populations, only Mramer is differentiated. It relatively has a higher number of unique and rare fragments, as well as relatively higher genetic diversity (46.12%). This population can be a promoter for research studies of specific genes of tolerance and adaptation and should have priority for conservation. Rbai and Admin populations can also be concerned since they have good genetic diversity (47.95% and 48.85% respectively). This study shows that AFLP markers can be considered as a powerful method to evaluate genetic diversity and estimate population differentiation of the argan tree. For more structuring feature, AFLP analysis should be combined with other numerical ones.

Materials and Methods

Plant materials and DNA isolation

In the present investigation, 130 individuals were sampled during the beginning of autumn from 13 populations (Table 4) of Essaouira different region following a random sampling

strategy. Young and healthy leaf samples were collected from adult trees, brought back to the laboratory, and preserved at -80°C until used. Total genomic DNA was extracted according to the ISOLATE II Plant DNA Kit. The quality and quantity of the DNA were determined using a NanoDrop 2000 (NanoDrop Technologies Inc., USA). They were also estimated by visual assessment on a 1% agarose gel.

AFLP analysis

AFLP analysis was carried out according to the procedures described by Vos et al. (1995). They recommended that the PCR amplifications of DNA fragments should be performed in two consecutive reactions.

The genomic DNA digestion and the adaptor ligation were performed following the protocol of AFLP Core Reagent Kit (Invitrogen, Carlsbad, CA). Genomic DNA (100-500 ng) was restricted with an enzyme mix of *EcoRI/MseI* (1.25 U/ μl each) in a restriction buffer (50 mM Tris-HCl pH 7.5, 50 mM Mg acetate, 250 mM K acetate) in a total volume of 25 μl during 2 hours at 37°C . After enzyme inactivation at 70°C for 15 min, *EcoRI/MseI* adaptors were ligated to DNA digested fragments using 1 U of T4 DNA ligase at 37°C for 2 h, to generate target sites for primer annealing and amplification. The first PCR reaction, called preselective amplification, was performed with primers complementary to the *EcoRI* and *MseI* adaptors with an additional selective 3' nucleotide. Since no work has been done on the argan genome size before, we used two specific kits, one specific to the small genome with the Arabidopsis DNA as a witness and the other specific to the regular genome using the Tomato DNA as a witness. The PCR conditions were as follows: 2 min at 72°C ; 20 cycles of 20 s at 94°C (denaturation), 30 s at 56°C (annealing) then 2 min of extension at 72°C . The amplification was confirmed by electrophoresis of PCR products on 1.5% agarose gel. The selective amplification was performed with 10 fold diluted pre-selective amplified PCR product in a total volume of 20 μl in two reactions using two sets of AFLP primers (Table 5). An *EcoRI* with two (for small genome) and three (for regular genome) selective nucleotides, at the 3' end and a fluorescent dyes in the 5' end, are used in combination with an unlabelled *MseI* primer with

three selective nucleotides. The EcoRI primers are labeled with FAM (blue), JOE (green), or NED (yellow).

PCR cycles started with an initial denaturation step at 94 °C for 2 min followed by 10 cycles at 94 °C for 20 s, 66 °C for 30 s, and 72 °C for 2 min. At each cycle the annealing temperature was decreased by 1 °C. The initial cycles were followed by 20 cycles at 94 °C for 30s, 56 °C for 30s, and 72 °C for 2 min. A final step at 60 °C for 30 min was added. All the polymerase chain reactions (PCR) were performed using a VERITI® Thermocycler (Applied biosystems). The selective PCR amplification was confirmed by electrophoresis as mentioned above. 1.5 µl of amplification products was mixed with 1.5 µl of deionized formamide and 0.15 µl GeneScan™ 500 ROX® as internal size standard (Applied Biosystems, Foster City, USA). These PCR products were denatured for 5 min at 95 °C, then quickly chilled on ice for another 5 min and loaded on an ABI 3130xl DNA analyzer for electrophoresis.

We performed primer combinations screening, on a subset of 13 samples by comparing their binary matrix and their replicates to detect unreliable markers. Four primer combinations (PC) [E-ACC/M-CTC (yellow), E-ACC/M-CTA (yellow), E-ACC/M-CAT (yellow), E-ACC/M-CAG (yellow)] (Table 1) were selected for further analysis since they generate various and distinct polymorphic loci.

AFLP data acquisition and statistical analysis

The AFLP data were scored using GeneMapper v5.0 software (Applied Biosystems, Foster City, CA, USA). The size of markers, locus selection threshold and phenotype calling threshold were considered as important scoring parameters. In our analysis, we included peaks within the size range of 70-500 bp. The data were scored as 'present' (1) or 'absent' (0) with threshold value of 100 relative fluorescence unit (rfu) RFU=100 as recommended by the ABI Plant fingerprinting manual), and a bin width ranging from 1 bp to 1.5 bp. GeneMapper automatically generated the binary matrix according to the parameter settings. The matrix of each AFLP profile was checked visually to remove uncertain or artifact peaks.

As the AFLP technique was never applied to *A. spinosa* L. before, an assessment of AFLP marker reliability and quality control was realized to estimate the error rate of genotyping. The AFLP profiles obtained from 13 replicated plant samples from different populations and all steps were re-analyzed from DNA extraction to allele scoring (Bonin et al., 2004). The error rate was computed as the ratio between observed number of phenotypic differences and total number of phenotypic comparison which can be assessed by comparing the binary matrix of 13 samples and their replicates (Bonin et al., 2004).

The genotypic data were used to calculate different parameters. The polymorphic information content (PIC) for each primer combination was calculated according to Roldán-Ruiz et al. (2000) formula: $PIC_i = 2f_i(1-f_i)$, where PIC_i is the polymorphic information content of marker i , f_i is the frequency of the present fragments. The term $1-f_i$ is the frequency of the absent fragments. PIC was averaged over the fragments for each primer combination. The marker index (MI) was calculated as proposed by Powell et al. (1996) as: $MI = PIC \times EMR$, where EMR (effective multiple ratio, $EMR = \beta \times n$) is defined as the product of the fraction of polymorphic loci (β) and the number of polymorphic loci (n). The resolving power (RP) of each primer combination was calculated following Prevost and Wilkinson (1999) as: $RP = \sum Ib$, where Ib is the fragment informativeness and

calculated as: $Ib = 1 - [2 \times |0.5 - p|]$, where p is the proportion of the genotypes containing the fragment. Genetic diversity and genetic structure were analyzed using different algorithms and methodologies. The genetic diversity parameters were calculated using the software GenALEX 6.5, such as the percentage of polymorphic loci (PPL), the number of different alleles (N_a), the number of effective alleles (N_e), the Nei's gene diversity index (H) and the Shannon information index (I) (Peakall and Smouse, 2006; Peakall and Smouse, 2012). The coefficient of gene differentiation (G_{st}), and gene flow (N_m) were calculated using POPGENE version 1.32 (Yeh et al., 1999). Analysis of Molecular Variance (AMOVA) was executed to estimate the distribution of genetic variation among and within populations using GenALEX 6.5 (Excoffier et al., 1992; Peakall and Smouse, 2006).

Finally, we conducted two analyses to visualize the distribution of individuals. First, the principal Coordinate Analysis (PCoA) was performed via a distance matrix to describe the relationship between individuals in populations using GenALEX (version 2013, Addinsoft™). Second, a dendrogram was also constructed according to a similarity matrix based on Jaccard's similarity coefficient depicting clustering pattern of individuals, using the Unweighted Pair Group Method with Arithmetic Means (UPGMA). A UPGMA tree is implemented by the software NTSYS-PC version 2.02 (Exeter software, New York).

Furthermore, an estimate of possible genetic exchanges between 13 *A. spinosa* L. populations was performed with Bayesian models implemented in STRUCTURE v2.1 (Pritchard et al., 2000). This analysis also confirms the results obtained from UPGMA and PCoA. The STRUCTURE v2.1 can identify genetically homogeneous groups and the number of differentiated clusters without a priori group designation, based on their multi-locus genotypes. This program was run for the range of genetic clusters from $K=2$ to 15 with 20 repetitions. The length of burn-in and MCMC (Markov chain Monte Carlo) were set to 10000, based on the work of Evanno suggesting that this level is sufficient (Evanno et al., 2005). We determined the optimal value of K using Structure Harvester (Earl and vonHoldt, 2011) based on the second order statistics (ΔK) developed by Evanno et al. (2005).

Conclusion

An AFLP analysis has been implemented on 130 genotypes from different parts of Essaouira region using two sets of primer combinations, one specific for small genome and the other specific for regular genome. The regular genome primer combinations have generated 477 AFLP loci. The statistical analysis differentiated the Mramer population from other populations. Thus, the Mramer population could be an interesting research subject for specific genes of tolerance and adaptation which should have the priority for conservation. The present results confirm that the AFLP technique, considered as a powerful technique, has revealed the structuring and the genetic differentiation of populations. This data can be explored in the context of rare tree conservation.

Acknowledgments

This work was realized in the framework of conducted in the project ArganBiogen and financially supported by Hassan II Academy of Science and Technology (Morocco) and The Ministry of Higher Education, Scientific Research and Professional Training of Morocco.

The authors express their deep gratitude to Dr. Adil Belhaj (University Sultan Moulay Sliman) and Mr. Soufian Ait Haddou (Consultant) for the linguistics review of this paper.

References

- Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor Appl Genet.* 102:222-229.
- Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. *Afr J Biotechnol.* 2:334-340.
- Alaoui K, Lagorce J-F, Cherrah Y, Hassar M, Amarouch H, Roquebert J (1998) Activité analgésique et anti-inflammatoire des saponines d'*Argania spinosa*. *Ann Pharm Fr.* 56:220-228.
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome Natl Res Counc Can Génome Cons Natl Rech Can.* 36:181-186.
- Aradhya KM, Phillips VD (1993) Genetic variability in fourteen provenances of *Eucalyptus* species in Hawaii. *Silvae Genet.* 42:9-15.
- Balasaravanan T, Pius PK, Raj Kumar R, Muraleedharan N, Shasany AK (2003) Genetic diversity among south Indian tea germplasm (*Camellia sinensis*, *C. assamica* and *C. assamica* spp. *lasiocalyx*) using AFLP markers. *Plant Sci.* 165:365-372.
- Bani-Aameur F, Ferradous A, Dupuis P (1999) Typology of *Argania spinosa* (Sapotaceae) fruits and stones. *For Genet.* 6(1):213-219.
- Baraket G, Chatti K, Saddoud O, Mars M, Marrakchi M, Trifi M, Salhi-Hannachi A (2009) Genetic analysis of Tunisian fig (*Ficus carica* L.) cultivars using amplified fragment length polymorphism (AFLP) markers. *Sci Hortic.* 120:487-492.
- Bechir A, Mousadik AE, Pichot C (2004) Allozyme diversity of natural and introduced cypress stands in Morocco: consequences for management of genetic resources. *Ann For Sci.* 61:8.
- Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals?. *Mol Ecol.* 14:2899-2914.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Mol Ecol.* 13:3261-3273.
- Charrouf Z, Guillaume D (1999) Ethnoeconomical, ethnomedical, and phytochemical study of *Argania spinosa* (L.) Skeels. *J Ethnopharmacol.* 67:7-14.
- Charrouf Z, Guillaume D, Driouich A (2002) L'arganier : Un atout pour le Maroc. *Biofutur.* 220:54-57.
- Chimi H, Cillard J, Cillard P (1994) Autoxydation de l'huile d'argan *Argania spinosa* L. du Maroc. *Sci Aliments.* 14:117-124.
- Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 4:359-361.
- Ait Aabd N, Msanda F, El MousadiK A (2015) Genetic diversity of the endangered argan tree (*Argania spinosa* L.) (sapotaceae) revealed by ISSR analysis. *J Agric Sci Rev.* 4:176-186.
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet.* 92:832-839.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol.* 14:2611-2620.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 131:479-491.
- Fay MF, Cowan RS, Leitch IJ (2005) The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Ann Bot.* 95:237-246.
- Fernández E, Figueiras M, Benito C (2002) The use of ISSR and RAPD markers for detecting DNA polymorphism, genotype identification and genetic diversity among barley cultivars with known origin. *Theor Appl Genet.* 104:845-851.
- Futuyma DJ (1986) Reflections on reflections: ecology and evolutionary biology. *J Hist Biol.* 19:303-312.
- Gupta A, Maurya R, Roy RK, Sawant SV, Yadav HK (2013) AFLP based genetic relationship and population structure analysis of *Canna*—An ornamental plant. *Sci Hortic.* 154:1-7.
- Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. AHD Brown, MT Clegg, AL Kahler, BS Weir (ed) *Plant population genetics breeding and genetic resources*, Sinauer Associates, Sunderland, MA. pp. 43-63.
- Hamrick JL, Godt MJW (1996) Conservation genetics of endemic plant species. In: *Conservation genetics, Case Histories from Nature* JC Avise and JL Hamrick (ed) Chapman and Hall, New York, pp. 281-304.
- Harhar H, Gharby S, Kartah B, Pioch D, Guillaume D, Charrouf Z (2014) Effect of harvest date of *Argania spinosa* fruits on Argan oil quality. *Ind Crops Prod.* 56:156-159.
- Hogbin PM, Peakall R (1999) Evaluation of the contribution of genetic research to the management of the endangered plant *ziera prostrata*. *Conserv Biol.* 13:514-522.
- Lamote V, Roldán-Ruiz I, Coart E, Loose MD, Bockstaele EV (2002) A study of genetic variation in *Iris pseudacorus* populations using amplified fragment length polymorphisms (AFLPs). *Aquat Bot.* 73:19-31.
- Lemos RPM, D'Oliveira CB, Stefenon VM (2015) Genetic structure and internal gene flow in populations of *Schinus molle* (Anacardiaceae) in the Brazilian Pampa. *Tree Genet Genomes.* 11:1-11.
- Li HY, Ru GX, Zhang J, Lu YY (2014) Genetic diversity in wild populations of *Paulownia fortune*. *Genetika.* 50:1330-1337.
- Lybbert TJ, Magnan N, Aboudrare A (2010) Household and local forest impacts of Morocco's argan oil bonanza. *Environ Dev Econ.* 15:439-464.
- Majourhat K, Jabbar Y, Hafidi A, Martínez-Gómez P (2008) Molecular characterization and genetic relationships among most common identified morphotypes of critically endangered rare Moroccan species *Argania spinosa* (Sapotaceae) using RAPD and SSR markers. *Ann For Sci.* 65:805-805.
- Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Mol Ecol.* 11:1145-1156.
- Mhirit O, Benziane M, Bencheikroun F, El Yousfi SM, Bendaanoun M (1998) L'arganier, une espèce fruitière forestière à usages multiples. *Mardaga, Sprimont, Belgique*, pp. 145-150.

- Msanda F, Peltier J-P, El Aboudi A (2007) Ecologie de l'arganier et typologie des arganeraies. In: Atlas de l'Arganier et de l'Arganeraie. In: IAV Hassan II (ed), Rabat, Morocco, pp. 87-107.
- Mueller null, Wolfenbarger null (1999) AFLP genotyping and fingerprinting. *Trends Ecol Evol.* 14:389-394.
- Naggar M, Mhirit O (2006) L'arganeraie : un parcours typique des zones arides et semi-arides marocaines. *Sci Chang Planétaires Sécher.* 17:314-317.
- Namkoong G, Boyle TJB, El-Kassaby YA, Palmberg-Lerche C, Eriksson G, Gregorius H-R, Joly H, Kremer A, Savolainen O, Wickneswari R, Young A, Zeh-Nlo M, Prabhu R (2002) Criteria and indicators for sustainability forest management: assessment and monitoring of genetic variation. *Cent Int For Res.* p 29.
- Ni X, Huang Y, Wu L, Zhou R, Deng S, Wu D, Wang B, Su G, Tang T, Shi S (2006) Genetic diversity of the endangered Chinese endemic herb *Primulina tabacum* (Gesneriaceae) revealed by amplified fragment length polymorphism (AFLP). *Genetica.* 127:177-183.
- Nouaim R (2005) L'arganier au Maroc: Entre mythes et réalités--Une civilisation née d'un arbre. L'Harmattan (ed), Paris.
- Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes.* 6:288-295.
- Peakall R, Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinforma Oxf Engl.* 28:2537-2539.
- Pecina-Quintero V, Anaya-López JL, Núñez-Colín CA, Zamarripa-Colmenero A, Montes-García N, Solís-Bonilla JL, Aguilar-Rangel MR (2013) Assessing the genetic diversity of castor bean from Chiapas, México using SSR and AFLP markers. *Ind Crops Prod.* 41:134-143.
- Polain de Waroux Y le, Lambin EF (2012) Monitoring degradation in arid and semi-arid forests and woodlands: The case of the argan woodlands (Morocco). *Appl Geogr.* 32:777-786.
- Pometti CL, Bessega CF, Vilardi JC, Cialdella AM, Saidman BO (2015) Genetic diversity within and among two Argentinean and one Mexican species of *Acacia* (Fabaceae). *Bot J Linn Soc.* 177:593-606.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed.* 2:225-238.
- Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet.* 98:107-112.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics.* 155:945-959.
- Roldán-Ruiz I, Dendauw J, Bockstaele EV, Depicker A, Loose MD (2000) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Mol Breed.* 6:125-134.
- Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénès C, Burg K, Kremer A (2004) Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics.* 168:1615-1626.
- Simental-Rodríguez SL, Quiñones-Pérez CZ, Moya D, Hernández-Tecles E, López-Sánchez CA, Wehenkel C (2014) The relationship between species diversity and genetic structure in the rare *Picea chihuahuana* tree species community, Mexico. *PLoS One.* 9(11):e111623.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science.* 236:787-792.
- Sparg SG, Light ME, Staden J van (2004) Biological activities and distribution of plant saponins. *J Ethnopharmacol.* 94:219-243.
- Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, Hoisington DA, Devi P, Varshney RK (2009) AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant. *Plant Sci.* 176:505-513.
- Vos P, Hogers R, Bleeker M, Reijans M, Lee T van de, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414.
- Wright S (1951) The genetical structure of populations. *Ann Eugen.* 15:323-354.
- Wu FQ, Shen SK, Zhang XJ, Wang YH, Sun WB (2014) Genetic diversity and population structure of an extremely endangered species: the world's largest *Rhododendron*. *AoB Plants.* 7:plu082.
- Yeh FC, Yang RC, Boyle TBJ, Ye ZH, and Mao JX (1999) POPGENE version 1.32. the user-friendly shareware for population genetic analysis. Molecular biology and biotechnology centre, University of Alberta, Edmonton, Canada.