

IDENTIFICATION OF *JUGLANS REGIA* L. GENOTYPES USING RAPDs MOLECULAR MARKERS

M. Abuín, R. Díaz & J. Fernández-López

Departamento de Producción Forestal, CIFA Lourizán, Apto. Correos 127, E-36080 Pontevedra, Spain

Received May 28, 2001; accepted March 12, 2002

ABSTRACT

In this work we performed a genetic identification of 56 clonal of walnut plus trees using the RAPD technique. The prescreening of 80 primers on three walnut genotypes allowed to select 8 primers which revealed polymorphism and showed experimental reproducibility. Fourteen variable bands were selected as RAPD markers from the total bands obtained. A cluster dendrogram and a principal component analysis was realized on the similarity data. The results obtained with both analysis have revealed that the genetic variability observed in the walnut genotypes was not structured into six localities of northwest of Spain.

Key words: walnut, RAPD, PCR, genetic identification

INTRODUCTION

Juglans regia (Persian walnut) is a native species of the mountain ranges of central Asia. In western Europe these trees have a scattered distribution, being mainly concentrated around human settlements, roads and crop fields whereas in eastern Europe they are also present as minor component of mixed forests. One of the main characteristics of the species is the multiple use, walnut is planted for producing both nuts and timber, in fact, walnut timber is considered one of the most valuable timber from European ones. This fact has caused an increment of the distribution area of the species and that the use of *Juglans regia* in Europe are becoming more important.

In these sense, CIFA of Lourizán has developed a breeding program focused on obtaining plants with good quality for forest use. Nowadays we have a germplasm bank for long term breeding with walnut selected plus trees and two progeny tests which are being evaluated in order to subsequently install a clonal seed orchard. Future objective in this work will be the vegetative propagation of best individuals. It is therefore very important to identify surely the original plus trees and for that it is necessary to apply methods for genetic identification.

Germplasm characterization is often based on morphological and agronomic traits which are usually affected by environment. In the last years, the use of molecular DNA markers, such as Randomly Amplified

Polymorphic DNA (RAPD) genetic markers generated by the polymerase chain reaction (PCR) using arbitrary primers has provided a new tool for the detection of DNA polymorphisms (WILLIAMS *et al.* 1990; WELSH & MCCLELLAND 1990). The RAPD marker technique is quick, easy and requires no prior sequence information, RAPD markers typically segregate as single (dominant) loci that can be scored for the presence or absence of a specific amplified DNA fragment (WILLIAMS *et al.* 1990). RAPD analysis have already proven to be valuable in genotype characterization as well as in population and pedigree analysis in a number of tree crop species, including apple (DUNEMAN *et al.* 1994; HARADA *et al.* 1993; KOLLER *et al.* 1993; MULCAHY *et al.* 1993), pistachio (HORMAZA *et al.* 1994), olive (BOGANI *et al.* 1994; FABBRI *et al.* 1995; WIESSMAN *et al.* 1995), *Citrus* spp (SANCHEZ-ESCRIBANO *et al.* 1996), *Prunus* spp. (CASAS *et al.* 1999), vine (LOUREIRO *et al.* 1998, TESSIER *et al.* 1999), and *Juglans* species (NICESE *et al.* 1997; 1998; MALVOLI *et al.* 1997).

In this work the application of RAPD molecular markers in walnut breeding program was used with two major goals: (1) to fingerprint walnut plus trees genotypes included in the germplasm bank of CIFA Lourizán and (2) to study the genetic diversity of the available germplasm.

Table 1. List of the walnut plus trees and their localities used in the RAPD analysis.

Tree	Code	Latitude	Longitude	Altitude	Tree	Code	Latitude	Longitude	Altitude
JRCO01 I	1	43.20	8.35	165	JROR15 S	29	42.42	7.78	165
JRCO02 II	2	43.36	8.18	65	JROR16 S	30	42.44	7.75	100
JRCO03 S	3	43.40	8.17	105	JRPO01 S	31	42.69	8.18	440
JRCO04 S	4	43.30	8.20	55	JRPO02 S	32	42.79	8.14	450
JRLU01 II	5	43.00	7.54	420	JRPO03 I	33	42.80	8.12	330
JRLU02 I	6	42.91	7.31	610	JRPO04 I	34	42.80	8.09	370
JRLU04 S	7	42.82	7.26	765	JRPO05 I	35	42.80	8.11	320
JRLU05 S	8	42.58	7.73	360	JROV01 I	36	43.07	6.53	750
JRLU06 S	9	42.90	7.27	540	JROV02 S	37	43.08	6.46	630
JRLU07 S	10	42.81	7.11	490	JROV05 I	38	43.24	6.35	400
JRLU08 S	11	42.80	7.14	725	JROV06 S	39	43.36	6.27	130
JRLU11 S	12	42.75	7.22	765	JROV07	40	43.35	6.18	90
JRLU12 S	13	42.76	7.23	700	JROV09 I	41	43.16	6.10	40
JRLU13 I	14	42.79	7.22	800	JROV10 I	42	43.43	5.61	410
JRLU13 PRS	15	42.50	7.72	400	JROV12 S	43	43.44	5.62	330
JRLU14 S	16	42.53	7.71	290	JROV13 S	44	43.40	4.98	220
JRLU15 S	17	42.54	7.71	320	JROV17 II	45	43.36	5.20	110
JRLU16 S	18	42.56	7.71	445	JROV18 I	46	43.35	5.16	220
JROR01 III	19	42.32	7.24	650	JROV18 S	47	43.35	5.15	210
JROR02 II	20	41.91	7.31	750	JROV19 S	48	43.40	5.15	60
JROR04 I	21	41.92	7.22	840	JRLE01 I	49	42.44	5.52	782
JROR05 I	22	41.94	7.21	590	JRLE03 S	50	42.51	5.54	785
JROR06 I	23	42.45	6.94	520	JRLE05 S	51	42.62	5.53	836
JROR08 I	24	42.46	6.90	400	JRLE06 S	52	42.67	5.63	865
JROR09 S	25	42.39	7.45	360	JRLE07 I	53	42.46	6.52	780
JROR10 S	26	42.38	7.50	250	JRLE09 S	54	42.49	6.58	252
JROR11 S	27	42.38	7.52	370	JRLE10 I	55	42.52	6.68	485
JROR13 S	28	42.38	7.54	430	JRLE12 S	56	42.71	6.92	680

MATERIAL AND METHODS

Plant materials

The 56 genotypes used in this study were collected at CIFA Lourizán walnut germplasm bank. The genotypes were from a selection of plus trees realized in six different localities of Northwest of Spain: Coruña, Lugo, Orense, Pontevedra, León and Oviedo (Table 1). Plus trees were selected by a comparison method based on phenotypic traits, grafted and planted in a germplasm bank. Two progeny tests from these plus trees have been included in 1998 and 1999. Clonal selection for wood production is being considered in the evaluation of the progeny test. Clonal selection both for wood and fruit will be further considered in the germplasm bank.

Genomic DNA isolation

Total DNA was isolated from fresh young leaves of these materials by the modified CTAB (cetyltrimethyl-ethyl ammonium bromide) protocol of LEFORT & DOUGLAS (1999). One hundred milligrams of fresh plant

material (leaf) were added to 0.750 ml of 65 °C pre-heated CTAB extraction buffer (50 mM Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, 0.7 M NaCl, 0.4 M LiCl, 1 % CTAB, 1 % PVP40, 2 % SDS and 2 % β -mercapto-ethanol) and incubated for 15 min at 65 °C. The mixture was emulsified with 0.5 ml of chloroform/isoamyl-alcohol (24:1), centrifugated for 5 min at room temperature and the aqueous layer retained. An equivalent volume of isopropanol was added to precipitate the nucleic acids. The DNA pellet was washed with 1 ml 70 % ethanol, air dried and finally DNA pellets were resuspended in 50 μ l of TE pH 8.0. DNA concentration was determined by the minigel method (SAMBROOK *et al.* 1989) by comparing with standard lambda DNA, and that diluted to approximately 30 ng/ μ l with TE.

DNA amplification

Amplifications were conducted with 10-mer primers from Operon Technologies Inc. (Alameda, Calif. USA). For the RAPD reactions, 30 ng of DNA were used as template in a final volume of 12.5 ml containing 1x reaction buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl,

0.1 mM EDTA, 1mM DTT, 50 % glycerol, 0.5 % Tween 20 and 0.5 % Nonidet P40), 3.6 mM MgCl₂, 0.156 mM of each dNTP (Ecogen), 0.57 mM primer and 1 U Taq polymerase (Ecogen). The mixture was overlaid with a drop of mineral oil (Sigma). The DNA amplifications were performed in a Perkin Elmer Gene Amp PCR System 9600 under the follow conditions: DNA denaturation was done at 94 °C for 3 min., followed by a 45 cycle amplifications (94 °C, 30 sec; 35 °C, 1 min.; 72 °C, 30 sec.) and by a final extension step at 72 °C for 6 min. Amplification products were analysed by electrophoresis in 1.5 % agarose gels in TAE buffer at 5 V.cm⁻¹ for 3 h. and detected by staining with ethidium bromide. The fragment patterns were photographed (667 Polaroid) under UV light for further analysis.

Data analysis

For each primer, the consistent amplified products were recorded from the examination of photographs. RAPD bands were scored manually as present (1) or absent (0). The polymorphic fragments (RAPD) were named by the primer code (OP Operon), the kit letter and its approximate size in base pairs.

Pair-wise comparisons of samples were used to estimate Jaccard's similarity coefficients (GS): $a/(n-d)$ where a = number of positive coincidences, n = total sample size and d = number of negative coincidences (JACCARD 1908). Cluster analysis were performed using the unweighted pair group method with arithmetic averages (UPGMA) and a dendrogram was constructed. Principal component analysis was performed on walnut genotypes bands frequencies (0 or 1). The first two principal coordinate were used to describe the dispersion of the 56 walnut plus trees according to their band data. All analysis were performed using the SAS analysis system software (version 8.01).

RESULTS AND DISCUSSION

Selection of RAPD markers

The results of an initial screening of 80 arbitrary RAPD primers (corresponding to kits A, B, C, D from Operon Technologies (Alameda, California)) in 3 of the walnut genotypes were: 7 did not produce amplification products, 25 amplified poorly and the remaining 48 primers produced clear, intense bands following PCR amplification. From this initial screen, 12 primers were chosen based on the quantity and quality of their amplified fragments and were then used to survey all walnut plus trees. Despite the large number of primers screened, walnut trees were largely monomorphic for all clearly interpretable RAPD bands, although there were differences in their intensity. Less intense bands that were sometimes absent in some individuals, could represent a polymorphism but was interpreted to be an artefact of weak amplification rather than a polymorphism at that position. From this selection, four primers failed to show polymorphisms among the 56 walnut plus trees and as a result eight primers were finally selected for the study.

The profiles obtained using RAPDs were difficult to score because all of the bands present in a single run were not always reproduced in a second analysis and the band intensities also differed in many cases between replicates. So, only bands that gave a reproducible score in the duplicated experiment were included in the final analysis. Among the 65 bands generated by the 8 selected primers 51 were present among all the samples analysed and correspond to monomorphic bands. The remaining 14 variable bands were selected as RAPD markers (represent 21.54 % of total bands) (Table 2). The selected primers generated distinctive products in the range of 300 bp to 2kb. Bands outside this range were not considered.

Table 2. Nucleotide sequence of selected primers with the number of total bands, polymorphic bands and the range in bp for each primer.

Primer code	Sequence 5'-3'	No total bands	No polymorphic bands	Range (bp)
OPA-06	GGTCCCTGAC	8	2	500–2000
OPA-08	GTGACGTAGG	9	3	350–2000
OPA-09	GGGTAACGCC	5	1	350–700
OPA-12	TCGGCGATAG	8	1	300–1600
OPA-15	TTCCGAACCC	8	2	250–900
OPA-17	GACCGCTTGT	10	1	300–1100
OPA-18	AGGTGACCGT	11	2	375–1200
OPB-19	ACCCCGAAG	6	2	450–900

Polymorphism and genotype identification

The 8 primers selected provided 14 RAPD markers that were used to clearly distinguish among the individuals. The analysis of the 56 walnut genotypes using the 14 selected RAPD markers allowed us to distinguish 49 different banding patterns. The following groups of plus trees: 1,33; 3,23,26,40; 5,6; 21,52 and 16,53 were indistinguishable and had identical banding patterns within each group (100 % similarity) in the RAPD analysis. Therefore, these genotypes will have to be assayed with additional RAPD markers with the aim to be unequivocally identified.

RAPD analysis revealed low levels of polymorphism within the reference set of *Juglans regia* genotypes. Although the number of bands for each primer varied from 1 to 11 with an average of 8 bands per primer, each of the 8 selected primers produced only one, two or, in the case of primer OPA-08, three RAPD markers (Table 2). This number is low compared with those obtained in other species: in apple, for example, 14 markers were generated by a single primer (KOLLER *et al.* 1993) or in the case of *Rubus* spp with only 10 primers they obtained 372 RAPD markers (GRAHAM & MCNICOL 1995). So this fact was one of major factors limiting the average number of RAPD markers per primer to 1–3 in our study (Table 2). The level of polymorphism (21.54 %) observed was much lower than 52.6 % reported by FJELLSTROM & PARFIT (1994) using RFLP markers, or 46.9 % reported by MALVOLTI *et al.* (1997) with isozymes as molecular markers in *Juglans regia* species. Our observation of lower levels of RAPD variation than allozyme and RFLP variation in other studies may simply be due to different capacities of the different markers to detect genetic variability. Several factors must be considered in using RAPD markers for assessing genetic variation in diploid outcrossing species. The principal factor is the dominant nature of RAPD markers. If a locus is polymorphic, with the recessive allele present at low frequency, almost all copies of that allele will be carried in heterozygotes and will go undetected by the RAPD procedure, resulting in that locus being considered as monomorphic. This will lead to an underestimate of polymorphism, which could be the explanation for our results (LIU & FURNIER 1993).

The matrix calculated for all possible pairwise comparisons between accessions showed that the similarity coefficients varied from: 0.09 to 0.83 with an average value of 0.36. The probability of finding two individuals with the same pattern can be calculated as the mean similarity index to the power of the mean number of bands per genotype (NYBOM & HALL 1991). In our case the probability of finding the same pattern

in two individuals is $2.37 \cdot 10^{-4}$ (i.e. $0.36^{8 \cdot 17}$). Although the use of 14 independent RAPD markers would allow the theoretical differentiation of 16,384 banding patterns (i.e. 2^{14}), our results showed that even 14 RAPD markers were not sufficient to identify all the genotypes. This fact suggests a linkage of some bands and/or the absence of some band combinations resulting from sampling (KHADARI *et al.* 1995).

Primers also differed in their capacity to differentiate between individuals genotypes. The possibility to obtain a unique genotype when RAPD markers are used in genetic identification will depend of three parameters: number of bands analyzed, number of variants per band and the frequency of variants in the sample. Thus, the more the variant frequencies have central values the lower the probability that two different genotypes will be included in the same identically characterized group. In this sense, some of the primers were not very useful to reliably distinguish the various individuals, because they were present or absent in only a few plants. The primers which showed allele intermediate frequencies (0.26 to 0.75) were OPA91, OPA152, OPA171, OPA181, OPA182 and therefore they were the most informative for walnut genotype identification.

Genotype relationship

The relationship among individuals illustrated by the dendrogram in figure 1 showed the absence of a clear genotype grouping, although the dendrogram showed some trends, thus, one group, containing provenances from León and Oviedo appears to be the most distantly related to all the others (Fig. 1). In order to assess whether the clustering of genotypes based on RAPDs could be further resolved, a principal component analysis (PCA) was carried out on the similarity matrices. PCA of allelic genotypes among the plants showed that only 17 % of the variation could be accounted for by the first PC, 13 % by the second PC, 10 % by the third PC and only 80 % by the first 8 PCs. A plot of the first two components from the PC analysis for the 14 polymorphic bands (Fig. 2) showed a similar tendency to the cluster analysis, that is, divide most of accessions from León and Oviedo regions from the rest of walnut clones. But otherwise separates few members of this latter group into any obvious pattern.

In order to facilitate the presentation of variable data we have used a graphical method termed "bandmap" (POWELL *et al.* 1991). The bandmap (Fig. 3) examines the relationship between genotypes and RAPD-derived amplification products. The presence of an amplification product is represented by a filled box and the order of the genotypes is exactly that generated by the dendrogram, thus genotypes which share a common

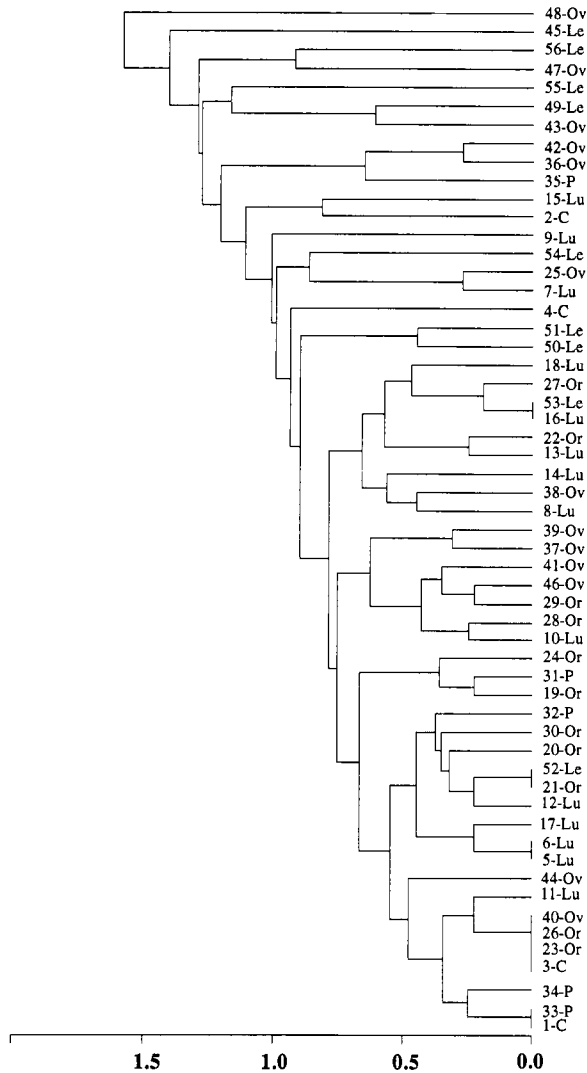


Figure 1. Dendrogram of walnut genotypes revealed by UPGMA cluster analysis of the Jaccard genetic similarity (GS) coefficient calculated from 14 RAPD amplification products. The numbers correspond to the code number of plus trees listed in table 1 and the abbreviations to: C: Coruña, Lu: Lugo, Le: León, Or: Orense, Ov: Oviedo and P: Pontevedra.

amplification product are more likely to be placed close to each other. The bandmap orders also the markers by their frequency. Figure 3 shows the more frequent RAPD markers placed on the top and the rare on the bottom. The plants showed a high number of shared markers, as illustrated in the bandmap and it was observed that some products, such as 19.1, 19.2, and 15.2 showed preferential absence in East provenances (León and Oviedo). These products could be therefore related with geographic provenances of the clones. It could be suggested that the absence of distinct group for the walnut genotypes studied here can be ascribed to

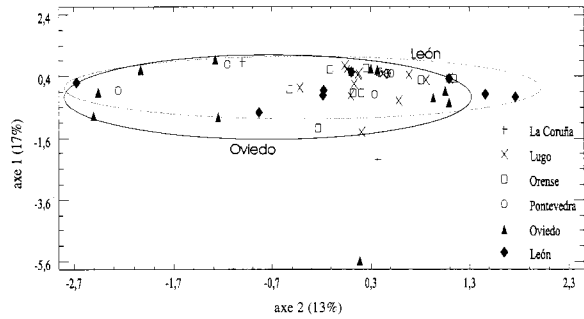


Figure 2: PCA plot from RAPD data set of 14 polymorphic bands from 8 primers. The two first axes of the PCA explain respectively 17 % and 13 % of the total variation.

different factors such as the small sample size or/and to the agricultural trade.

CONCLUSIONS

This study showed that the RAPD technique can be used to generate amplified segments of genomic DNA that can differentiate walnut genotypes. Since pedigree data are unknown for the walnut plus trees studied, RAPD markers have showed to be very useful to assess the degree of similarity of genotypes in this woody species, in order to select the best parents to obtain new genetic combinations in a breeding programme. This is especially important if consider the long generation times of nut tree species and consequently the long time of the breeding process. Given the results so far it should be possible to establish a fingerprint reference library with the objective to check the identity of propagated clones. The library will serve as a reference source of RAPD profiles for each clone to allow them to be identified.

ACKNOWLEDGEMENTS

This study was financed by the Instituto Nacional de Investigaciones Agrarias (INIA) from the Spanish Ministry of Agriculture, Science and Technology (SC99–SCOO-006 project). M. Abuín is a recipient of a postdoctoral fellowship from the Spanish Ministry of Agriculture, Science and Technology.

REFERENCES

BOGANI, P., CAVALIERI, D., PETRUCELLI, R., POLSINELLI, L. & ROSELLI, G. 1994: Identification of olive tree cultivars by using Random Amplified Polymorphic DNA. *Acta Horticulturae* **356**: 98–101.
CASAS, A. M., IGARTUA, E., BALAGUER, G. & MORENO, M. A.

- WIESMAN, Z., AVIDAN, N., LAVEE, S. & QUEBEDEAUX, B. 1998: Molecular characterization of common olive varieties in Israel and the West Bank using randomly amplified polymorphic DNA (RAPD) markers. *Amer. Soc. Hort. Sci.* **123**(5): 837–841
- WILLIAMS, J. G. K., KUBLELIK, A. R., LIVAK, K. J., RAFALSKI, J. A., TINGEY, S. V. 1990: DNA polymorphism's amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**:6531–6535.