VARIATION OF RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD) MARKERS IN PECAN

M. lmbalzano[1,2/] and M. Stine[1/]

Abstract.–The potential use of RAPD markers to measure intraspecific variation in pecan *(Carya illinoensis* (Wang.) K. Koch) was examined. Three parent/progeny lines were screened with nine Operon ® primers to identify RAPD markers and to calculate their inheritance ratios. These markers were then identified in 77 individuals from a provenance test. Based on chi-square analysis,

Mendelian fashion. Of the bands scored, 24% in the inheritance study and 33% in the provenance study were not reproducible. Phylogenetic analyses were conducted on the full data set of markers identified in the provenance test and three subsets of the data. The first subset contained only the heritable and reproducible markers, the second included only the heritable markers, and the third contained only the reproducible markers. For none of the data sets examined does the RAPD data agree with anticipated clustering based on geography, with the exception of a Louisiana and western Mississippi cluster. The potential causes of these results are discussed.

Keywords: *Carya illinoensis* (Wang.) K. Koch, polymerase chain reaction, molecular markers.

INTRODUCTION

The development of the RAPD technique by Williams et al. (1990) and the arbitrarily primed DNA technique by Welsh and McClelland (1990) have raised interest in the application of RAPDs in many areas of research. RAPDs can be less time-consuming and less expensive than other methods, such as restriction fragment length polymorphisms (RFLPs), which makes it very appealing to many researchers. Studies have demonstrated that RAPD primers can be

---

[1/] Graduate Research Assistant and Assistant Professor, Louisiana State University Agricultural Center, Louisiana Agricultural Experiment Station, School of Forestry, Wildlife, and Fisheries, Baton Rouge, LA 70803-6200

[2/]Current Address: The Worcester Foundation for Experimental Biology, Shrewsbury, MA 01545

used to identify DNA polymorphisms in a wide variety of animal, plant, and bacterial species (Welsh and McClelland 1990, Williams et al. 1990, Goodwin and Annis 1991, Hu and Quiros 1991).

Many potential applications for this technique are currently being explored. Its usefulness in the determination of taxonomic identity has already been demonstrated. Hu and Quiros (1991) have shown that RAPD markers can differentiate among cultivars of broccoli *(Brassica oleracea)* and among cultivars of cauliflower (B. *oleracea).* Arnold et al. (1991) demonstrated that RAPD primers could produce species-specific markers for irises, which were then used to demonstrate hybridization and introgression. Using six maize inbred lines and five hybrids from the inbreds, Welsh et al. (1991) conducted a double blind experiment that used RAPD markers to identify the parents of the hybrids. RAPD markers may also have potential use in phylogenetic studies. Halward et al. (1992) identified species-specific markers for 29 diploid wild species of peanut *(Arachis hypogaea).* The characters were analyzed by two phylogenetic reconstruction programs, PAUP and HyperRFLP, to determine species relationships. The two programs produced almost identical dendrograms, which were consistent with dendrograms produced with morphological, enzymatic, and RFLP data.

Based on the usefulness of RAPD markers as demonstrated in the literature, we wished to conduct a study on the potential use of RAPD analysis to investigate intraspecific variation in pecan *(Carya illinoensis* (Wang.) K. Koch). To determine if the RAPD characters would be useful in cladistic analysis, the data were used in the phylogenetic reconstruction program, Hennig86 (Farris 1988). A clustering of individuals according to their geographic source was expected.

MATERIALS AND METHODS

Leaf tissue was collected from 77 individuals in a pecan provenance test, located at Idlewild Experimental Station near Clinton, Louisiana. The provenance test contains accessions from 10 states throughout the natural range of pecan. To verify the inheritance of the RAPD markers, the leaf tissue of three pecan pedigrees, each consisting of two parents and 10 progeny, was obtained from the Pecan Genetics and Breeding Station Orchard in Brownwood, Texas.

A modification of Murray and Thompson's (1980) method was used to extract total cellular DNA from the tissue samples (5% polyvinylpyrrolidone (mw 40,000) was added to the extraction buffer). The DNA was further purified using the BioRad Prep-A-Gene" DNA Purification Kit. The RAPD protocol was a modification of the Williams et al. (1990) procedure (C. D. Nelson, pers. comm.). RAPD reactions were performed in 25 µI volume of 10x buffer (10 mM Tris-CI, pH 8.3; 50 mM KCI), 2.0 mM of $MgCl_2$, 100 µM each of dATP, dCTP, dGTP, and dTTP, 0.2 M of primer, 2.5 ng of template DNA, and 1.0 unit of *Taq* DNA polymerase and overlaid with 35 µl of mineral oil. A 96-well MJ Research Programmable Thermal Controller® set for 41 cycles of 1 min at 92 °C (denaturation), 1 min at 35 °C (primer annealing), and 2 min at 72 °C (primer extension) was used for amplifying the DNA. The amplified products were separated in a 1.4% agarose gel in 1x TAE buffer.

Based on the screening of 56 Operon primers (C1-20, F1-20, W1-16) on bulked samples of pecan DNA, 9 primers were selected to fingerprint the individuals in the provenance test and to verify the inheritance of the RAPD markers in the three pedigrees. In addition, 25% of the total number of reactions from the provenance test and 100% of the total reactions for the inheritance study were repeated in order to determine if the DNA banding patterns were reproducible. Negative controls, in which no template DNAs were added to the reaction mixtures, were run for each primer.

Polymorphic and monomorphic loci were scored as either present or absent. Any amplified products found in the pecan sample lanes that corresponded to products in the negative control lanes were not included in the data set. Also, any loci that were not scorable due to clarity for all the individuals in the provenance test were not included. Chi-square analysis was conducted on the loci of individuals from the Brownwood orchard to determine if the loci were inherited in a Mendelian fashion.

The provenance data set was divided into three subsets based on the heritability and the reproducibility of the DNA banding patterns. The first included only the loci that were inherited in a Mendelian fashion and were reproducible in both the inheritance study and in the provenance test. The second subset included all of the heritable loci, whether they were reproducible or not. The third subset consisted of all of the reproducible markers, whether or not they were inherited in a Mendelian fashion.

Cluster analysis of the full data set and the three subsets was performed using Hennig86 vers. 1.5 (Farris 1988), which was distributed to R. M. Zink by the program designer, J. S. Farris. Cladograms of each of the data sets were created using a heuristic search option due to the large number of individuals in the data sets. From the first 100 equally parsimonious trees, a consensus tree was formed for each data set and was used for analysis.

RESULTS

Out of the 56 primers screened, 10 produced no amplification or poor amplification, 20 amplified only monomorphic loci, and 26 amplified at least 1 polymorphic locus. From the 26 primers that produced polymorphic bands, 9 primers were selected to fingerprint the individuals from the Brownwood orchard and the provenance studies (C6, C12, F7, F14, W3, W4, W5, W8, and W16).

From the individuals in the inheritance study, a total of 58 loci, ranging from 0.4 kb. to 2.5 kb., were produced with the 9 primers. After repeating all of the RAPD reactions, 44 of the 58 markers were reproducible. Chi-square analysis was conducted on the presence:absence ratios of the individuals to determine if they were inherited in a Mendelian fashion. None of the observed ratios was rejected ($\alpha = 0.05$), with the exception of cases where a marker was scored in some or all of the progeny but not in either of the parents, which accounted for approximately 25% of the markers. Of the 58 loci, 44 were inherited in a Mendelian fashion, and 38 were heritable and reproducible.

From the individuals in the provenance test, a total of 63 loci, 47 polymorphic and 16 monomorphic, ranging from 0.4 kb. to 2.5 kb., were scored based on the 9 primers chosen. Of the 63 loci, 42 were reproducible based on a check of 25% of the total number of RAPD reactions, and 54 were also scored in the individuals from the inheritance study. Based on chi-square analysis of the presence:absence ratios of the loci scored in the inheritance study, 30 loci were found heritable and reproducible. Only 28 of these loci (16 polymorphic loci) were heritable and reproducible in both the inheritance and provenance studies, and 13 loci were found heritable and nonreproducible in at least one of the studies.

Due to the problems encountered with the reproducibility of the markers, we were interested in how reproducibility and heritability of the markers would affect the clustering of the data when analyzed by Hennig86. Therefore, the full data set of 63 markers (47 polymorphic) for the 77 individuals was divided into three subsets as described previously. The cladograms for the full data set and the heritable and reproducible data subset are shown in figures 1 & 2 respectively. The full data set produced a 356-step concensus tree with a consistancy index (CI) =0.13 and a retention index (RI) = 0.57. The heritable and reproducible data set produced a 41-step consensus tree with a CI =0.39 and RI =0.81. Both CI and RI can range from 0.0 to 1.0, with values closer to 1.0 indicating a close fit between the data and the tree. The full data set revealed more clustering than the heritable and reproducible data set, but none of the cladograms showed strong congruence to anticipated patterns based on geography.

DISCUSSION

The results of the Hennig86 analysis for all four data sets show only limited clustering of the individuals by geographic location, with the exception of a group of individuals from Louisiana and Mississippi. With the addition of the less reliable data (nonreproducible, nonheritable, or both), more clustering of individuals or populations was apparent, but the CI and RI values decreased. Not all of the clusters included samples from the same or nearby populations. In fact, in most cases, all of the individuals from the same population did not group together, which would indicate that there is a significant degree of variation within populations as well as among them.

The nature of the RAPD markers may also introduce sources of error into the phylogenies. One of the limitations of RAPDs is that it is not known what level of variation in the genome is being detected by the RAPD markers. For example, it cannot be easily determined if a RAPD marker represents the presence of a single copy gene or a multicopy gene. The absence of a marker representing a multicopy gene will only occur if all of the loci are not present. However, the number of present loci of a multicopy gene may be phylogenetically important. In this situation, using RAPD analysis may result in the underestimation of the actual amount and distribution of variation of RAPD markers.

In order to have the highest level of confidence in the information obtained from RAPD analysis, a complete genetic system of the study species, including several generations of parents, progeny, and backcross individuals, should be available. RAPD primers can then be screened on the individuals in the system and further analysis utilize markers that segregate in a Mendelian fashion. The limited availability of pedigrees to evaluate the RAPD markers limits

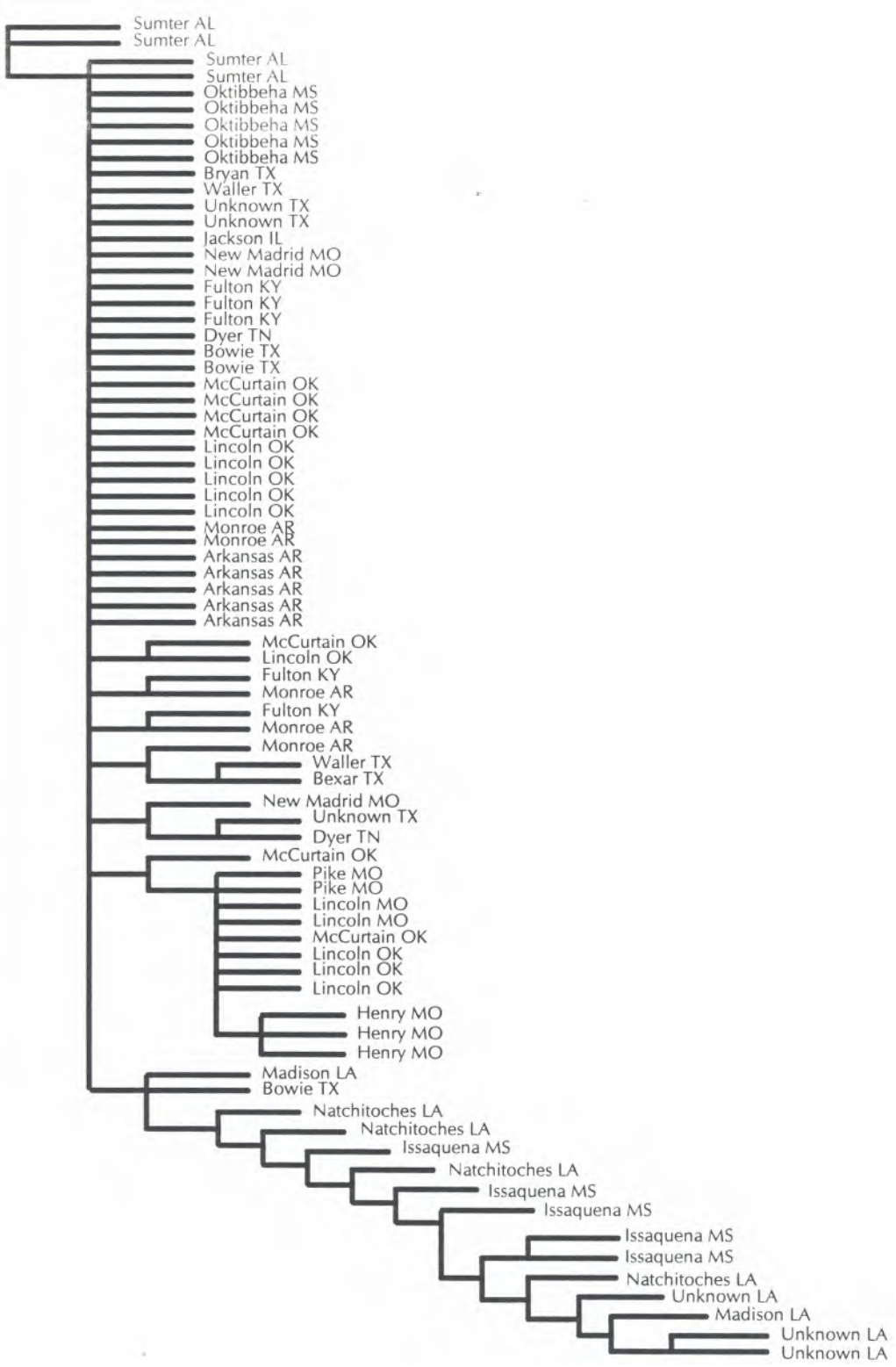

Figure 1. Cladogram generated by Hennig86 using all of the identified marker.

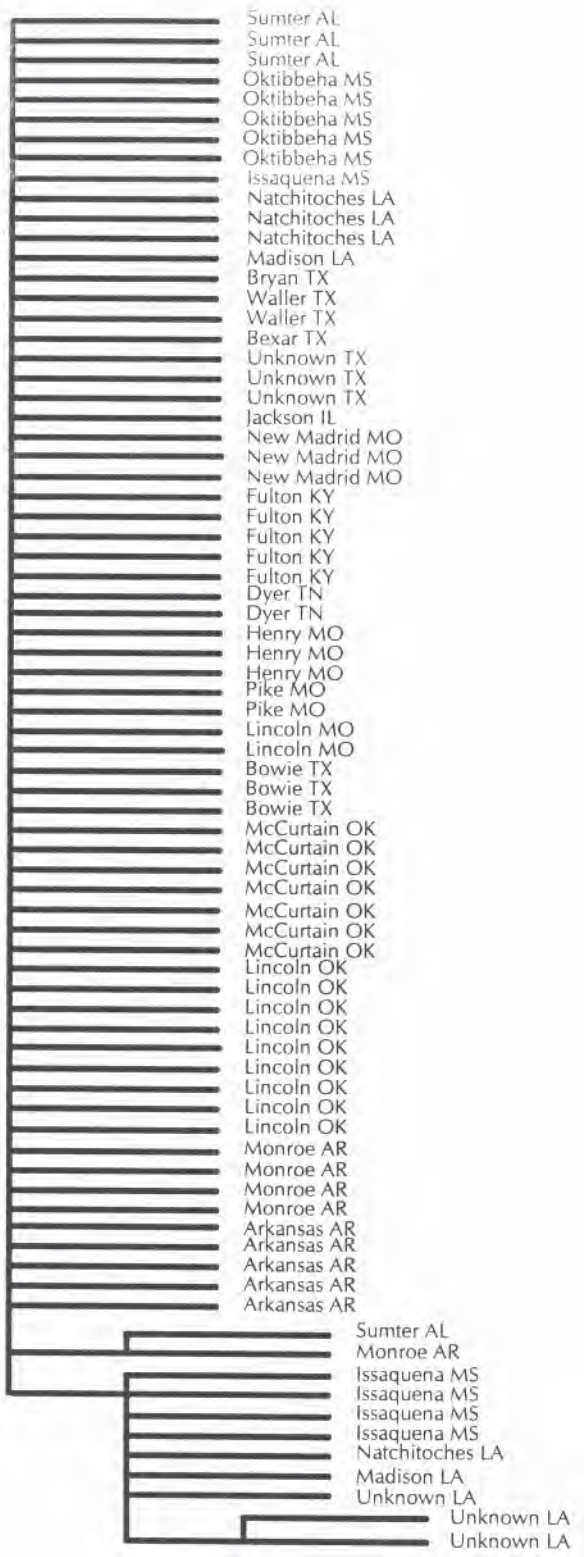

Figure 2. Cladogram generated by Hennig86 using only the heritable and reproducible markers.

the applicability of this technique in phylogenetic analysis and many studies involving wild populations. This may not be a major obstacle for plant species that have an annual life cycle; but for tree species, the long generation times and absence of large pedigrees for most species is a considerable problem.

For this study, only three pedigreed families were available to screen primers and select markers demonstrating Mendelian inheritance. Because of the small sample size, some markers that were not inherited in a Mendelian fashion may have been included in the data set. A larger number of progeny and having more than one generation would reduce this problem. Additionally, none of the three pedigreed lines were from individuals in the provenance test. While it may be acceptable to assume that the markers found in the inheritance study were comparable to those identified in the provenance test, having the genetic system for the individuals in the provenance test would increase the reliability of the results.

The problems of nonreproducible and nonheritable data are not the only possible reasons for the lack of clustering according to geographic location. The natural history of pecan provides additional theories for the degree of randomness demonstrated in the clustering of individuals. It is believed that pecan originated 70-135 million years ago during the Cretaceous period (Stuckey and Kyle 1925), with the present range of pecan determined since the last glaciations of the Quaternary period (Delcourt and Delcourt 1987). Pecan spread northward from the southern part of its present range primarily through the activities of humans (Bettis et al. 1990). Native Americans were primarily responsible for further expansion of the range of pecan during the 1 7th and 18th centuries (Hume 1906, Stuckey and Kyle 1925, Woodruff 1979). By the mid-1 700s, fur traders had introduced the pecan nut to English settlers in the eastern United States. By collecting pecans in one area and planting them in another, they succeeded in extending the range of pecan and in scattering genotypes into regions where they would not have normally been introduced. The spread of the natural range of pecan by humans could help explain the overall lack of meaningful clustering of individuals in the cladograms.

In conclusion, the development of the RAPD technique may provide a new DNA-based marker that can be obtained through relatively simple and inexpensive means. However, the reliability of these RAPD markers must be improved in order to obtain a higher level of confidence in some of the studies using the technique, particularly those involving phylogenetic analyses and wild populations. Modifying the protocol to increase the reproducibility of the markers and evaluating the inheritance of the markers through a complete genetic system consisting of several generations of parents and progeny are two possible methods of increasing the reliability of the RAPD technique.

ACKNOWLEDGEMENTS

We appreciate the efforts of M.S. Bowen, T.J. Dean, T.S. Kubisiak, and V.W. Wright for reviewing this manuscript. This work was supported in part by funds from McIntire-Stennis project 2895 and Louisiana Stimulus for Excellence in Research and National Science Foundation EPSCoR Program grant NSF(EPSCoR)/LaSER 2/92-2/96. Approved for publication by the Director of the Louisiana Agricultural Experiment Station as Manuscript Number 93-22-7198.

LITERATURE CITED

Arnold, M. L., C. M. Buckner, and J. J. Robinson. 1991. Pollen-mediated introgression and hybrid speciation in Louisiana irises. Proc. Natl. Acad. Sci. USA. 88:1398-1402.

Bettis III, E. A., R. G. Baker, B.K. Nations, and D. W. Benn. 1990. Early Holocene pecan, *Carya illinoensis,* in the Mississippi River Valley near Muscatine, Iowa. Quart. Res. 33:102-107.

Delcourt, P. A. and H. R. Delcourt. 1987. Long-Term Forest Dynamics of the Temperate Zone. ed. W. D. Billings, F. Golley, 0. L. Lange, J. S. Olson, H. Remmert. New York: Springer-Verlag. 439 pp.

Farris, J. S. 1988. Hennig86, version 1.5. Distributed by the author to R. M. Zink. Port Jefferson Station, NY.

Goodwin, P. H. and S. L. Annis. 1991. Rapid identification of genetic variation and pathotype of *Leptosphaeria maculans* by random amplified polymorphic DNA assay. Appl. Environ. Microbiol. 57:2482-2486.

Halward, T., T. Stalker, E. LaRue, and G. Kochert. 1992. Use of single-primer DNA amplifications in genetic studies of peanut *(Arachis hypogaea L.).* Plant Mol. Biol. 18:315-325.

Hu, J. and C. F. Quiros. 1991. Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep. 10:505-511.

Hume, H. H. 1906. Pecan and Its Culture. Harrisburg, PA: Mt. Pleasant Press. 195pp.

Murray, M. G. and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8:4321-4325.

Stuckey, H. P. and E. J. Kyle. 1925. Pecan-Growing. New York: MacMillan Co. 233 pp.

Welsh, J., R. J. Honeycutt, M. McClelland, and B. W. S. Sobral. 1991. Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR). Theor. Appl. Genet. 82:473-476.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:7213-7218.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Woodruff, J. G. 1979. Pecans. In Tree Nuts: Production, Processing, and Products. pp. 327-364. Westport, CT: Avi Publishing Company, Inc.