Theor Appl Genet (1993) 85:697–701

Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars*

J. I. Stiles[1], C. Lemme[1], S. Sondur[1,2], M. B. Morshidi[2], and R. Manshardt[2]

[1] Department of Plant Molecular Physiology, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI 96822, USA
[2] Department of Horticulture, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI 96822, USA

Received April 6, 1992; Accepted May 7, 1992
Communicated by G. Wenzel

Summary. We have applied the recently developed technique of random amplification of polymorphic DNA (RAPD) to the analysis of the relationships among ten cultivars of papaya (*Carica papaya* L.). Eleven ten-base synthetic oligonucleotides were chosen that gave multiple PCR amplification products using papaya DNA as template. These 11 primers amplified a total of 102 distinct fragments. Cultivars were scored for presence or absence of RAPD fragments and grouped by cluster analysis using simple matching coefficients of similarity. A dendrogram of the ten cultivars was constructed. Of the ten cultivars seven were of the Hawaiian type, and all of these grouped to one branch of the tree. Divisions within the Hawaiian, branch were mostly consistent with the known genetic background of these cultivars. Three non-Hawaiian, cultivars were also analyzed. The minimum similarity detected was 0.7 suggesting that the domesticated papaya germ plasm is quite narrow. Our results show that RAPD technology is a rapid, precise and sensitive technique for genomic analysis.

Key words: *Carica papaya* L. – Polymerase chain reaction – Genetic polymorphism – Phylogenetic relationship

Introduction

Recent advances in techniques for DNA analysis and subsequent data analysis have greatly increased our ability to understand the genetic relationships among organisms at the molecular level. From its first use to identify DNA sequence polymorphisms for genetic mapping in 1975 (Grodzicker et al. 1975), RFLP analysis has developed into a major new tool for genetics and breeding. This technology has not only allowed the rapid development of saturated genetic maps but has also made possible the dissection of quantitative traits into Mendelian factors (Paterson et al. 1988; Lander and Botstein 1989). More recently, a rapid and sensitive technique using the polymerase chain reaction (PCR) has been introduced (Williams et al. 1990; Welsh et al. 1991; Welsh and McClelland 1991). This technique uses short synthetic deoxyribonucleotides of random sequence as primers for PCR. While the PCR products are produced from random regions of the genome, they are specific and reproducible since they are primed from specific DNA sequences within the genome. The approach of Williams et al. (1990) uses 10-base synthetic oligonucleotides with a GC content of 50–60%. The number of potential 10-base primers is very large so that numerous polymorphisms between even closely related organisms can be obtained. A further advantage is that random amplification of polymorphic DNA (RAPD) is much faster that traditional RFLP analysis. RAPD data is obtained by staining agarose electrophoresis gels containing fragments synthesized in a few hours using the automated technology of PCR, as opposed to the cloning of fragments, Southern blotting and autoradiography of gels that is necessary for RFLP data acquisition. In addition to mapping studies, Welsh et al. (1991) have shown that this technique can be used for assessing parentage in maize hybrids.

We have used RAPD for an analysis of the relationships among ten different cultivars of papaya (*Carica papaya* L.). The relationships of some of these

* Journal series No. 3692 of the Hawaii Institute of Tropical Agriculture and Human Resources
Correspondence to: J. I. Stiles

cultivars is known in detail, while others are of more uncertain origin. Seven of the ten cultivars are based on, or contain, appreciable amounts of the Hawaiian germ plasm. These include Line 40, 'Kapoho', 'Waimanalo' (Nakasone et al. 1972), 'Sunset' and 'Sunrise' (Hamilton and Ito 1968). 'Exotica' is derived from a Malaysian variety 'subang-6' that was back-crossed several times to 'Sunrise' (Chan 1985; Chan 1987), while 'Kamiya' is an open-pollinated selection from 'Waimanalo'. Three other cultivars, one a dioecious line obtained from Florida, USA (courtesy of Dr. R. A. Conover) that is thought to be less highly inbred than the other cultivars used in this study (Line 356) and two gynodioecious types ('Saipan Red' from the Mariana Islands and 'Bentong' from Malaysia) were also analyzed. These three cultivars are not known to be related to each other or to the Hawaiian, lines.

Multivariate analysis produced a dendrogram that grouped cultivars of known pedigree largely as predicted from their genetic makeup, with two exceptions. The relationships among cultivars of less well-defined genetic background were inferred from our data. Based on our results we believe that RAPD will have considerable utility as a technique to rapidly estimate relationships between closely related and more distantly related species and cultivars.

Materials and methods

Growth of plants

All cultivars were field grown at the University of Hawaii Poamoho Research Farm, Oahu, Hawaii. One plant of each of the following inbred, gynodioecious cultivars was analyzed: Line 40, 'Waimanalo', 'Kamiya', 'Kapoho', 'Sunrise', 'Sunset', 'Exotica', 'Saipan Red', and 'Bentong'. Two plants of the dioecious Line 356, one staminate and the other pistillate, were analyzed.

Extraction of DNA

Young leaf tissue was excised from mature plants. After removal of the midrib and other large veins, the remaining tissue was cut into small pieces and frozen in liquid nitrogen. Frozen leaf tissue was stored at −20 °C. The grinding chamber of a standard coffee grinder (Krups 'Touch-Top Coffee Mill') was cooled with dry ice. About 10 g leaf tissue and a small amount of dry ice was placed in the grinder and homogenized for about 20 s. DNA was extracted using a modification of the method of Dellaporta et al. (1983). The powdered leaf tissue was transferred to a beaker, and 150 ml of extraction buffer [100 m*M* TRIS-HCl (pH 8.0), 50 m*M* EDTA, 500 m*M* NaCl, 30 m*M* β-mercapto-ethanol and 1.3% sodium dodecyl sulfate (SDS)], heated to 65 °C, was added. After 10 min at 65 °C, the cell debris was removed by centrifugation at 6870 *g*. The DNA was precipitated by the addition of 0.8 volume of propanol and recovered by centrifugation for 15 min at 6870 *g* after incubation on ice for at least 30 min. The pellet was dried and redissolved in 10 ml of TE [10 m*M* TRIS-HCl (pH 8.0), 0.5 m*M* EDTA]. NaCl was added to a final concentration of 0.7 *M*, and an equal volume of 1% cetyltrimethylammonium bromide (CTAB) was added. After 30 min at room temperature the precipitate was collected by centrifugation at 12 100 *g* for 10 min. The pellet was washed with 70% ethanol containing 0.3 *M* sodium acetate and redissolved in a minimum volume (about 2 ml) of TE containing 0.3 *M* sodium acetate. An equal volume of 4 *M* ammonium acetate was added. After 30 min at 4 °C, the sample was centrifuged for 10 min at 12 100 *g*. The DNA was collected by precipitation after the addition of 0.8 volume of propanol as before. The DNA was dissolved in 0.5 ml TE and made 0.3 *M* by the addition of 1/10 volume of 3 *M* sodium acetate. Contaminating RNA was removed by digestion with 10 µg RNase A for 30 min at 37 °C. The DNA was further purified by extracting twice with an equal volume of phenol followed by extraction with an equal volume of chloroform. The DNA was precipitated by the addition of 0.8 volume of propanol and recovered as before. The final pellet was dissolved in 0.5 ml TE and the DNA concentration was determined by fluorometry using a Hoefer DNA fluorometer and following procedures supplied by the manufacturer.

Amplification of DNA

PCR was carried out in 25 µl reactions containing 15 ng papaya DNA, 2 µ*M* primer, 150 µ*M* each deoxyribonucleotide triphosphate, 2 m*M* $MgCl_2$, 10 m*M* TRIS-HCl (pH 8.3), 50 m*M* KCl, 0.001% gelatin and 0.75 U Taq DNA polymerase (Perkin Elmer Cetus). Amplification was for 45 cycles, each consisting of 94 °C for 1 min, 35 °C for 1 min and 72 °C for 2 min. Products were analyzed by electrophoresis in 1.5% agarose gels at 20 V for 16–18 h, stained with ethidium bromide and photographed under UV light.

Data analysis

Eleven primers (Table 1) each ten bases in length were selected at random from a pool of primers that gave reasonable numbers of strong amplification products using DNA from 'Sunrise' as template and the PCR reaction conditions described above. Each cultivar was scored for the presence or absence of every amplification product and the data entered into a binary data matrix. Coefficients of similarity, either simple matching or Jaccard's, were calculated; cluster analysis, using the UPGMA method, was performed; and a phenogram was produced as described by Sneath and Sokal (1973) using the Numerical taxonomy and multivariate analysis system for the IBM PC, version 1.60 (Applied Biostatistics). Parsimony analysis was done using the PAUP computer program (Swofford 1990).

Table 1. Synthetic deoxyribonucleotides used as primers for amplification of papaya DNA

Primer number	Nucleotide sequence (5′ to 3′)	Total number of fragments amplified
1	ACCAGGCCAA	16
2	ATCTGTGTGG	6
3	CACAGGTTCT	7
4	CAGTCCTAGG	8
5	CTAAGCCATG	9
6	GCATCTCAGT	9
7	CTTGGGTTGG	6
8	AAAGGCAACC	8
9	GATCGACACT	14
10	ACGGACTGGA	7
11	GTGGCTCTGA	12

Results

Although most of the 102 PCR amplification products were common to all cultivars, sufficient polymorphism was present to distinguish every variety and to distinguish the two individual plants of Line 356.

An example of a typical analysis is shown in Fig. 1. Oligonucleotide 2 (Table 1) primed the amplification of six specific fragments, three of which are present in all cultivars and three of which are missing in different combinations from certain cultivars.

Similarity among cultivars was estimated using distance matrix and parsimony numerical taxonomic methods on the 102 amplification products scored. Among the comparison methods used, the simple matching coefficient gave relationships that were most consistent with the known pedigrees of these cultivars. Coefficients of similarity ranged from about 95% for the most closely related cultivars to about 70% for those more distantly related (Table 2). Cultivars known to contain large amounts of Hawaiian germ plasm had similarity coefficients above 78% (Table 2) and, as expected, grouped to a single branch of the tree (Fig. 2). Relationships among the Hawaiian cultivars were mostly as expected from pedigrees, with the exceptions of 'Sunset' and 'Kamiya'. 'Sunset' originated as a sib line from the same F_2 population as 'Sunrise'. The affinity of 'Sunset' to 'Waimanalo', indicated by the dendrogram (Fig. 2), was unexpected since the two cultivars have no common parent. On the other hand, the RAPD data indicated that 'Kamiya', an open-pollinated selection from 'Waimanalo', was more closely related to 'Sunrise' than to 'Waimanalo'. The cultivars not of Hawaiian lineage, Line 356, 'Saipan Red' and 'Bentong', grouped separately and converged at about the same point, 80% (Fig. 2).

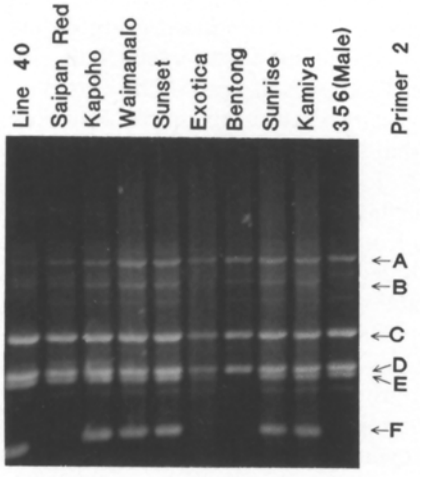

Fig. 1. Agarose gel of RAPD-amplified papaya DNA. DNA from each of the ten papaya cultivars was amplified using primer 2 (Table 1) and separated on a 1% agarose gel as described in the methods. Amplification fragments *B*, *E* and *F* show differences among the ten cultivars

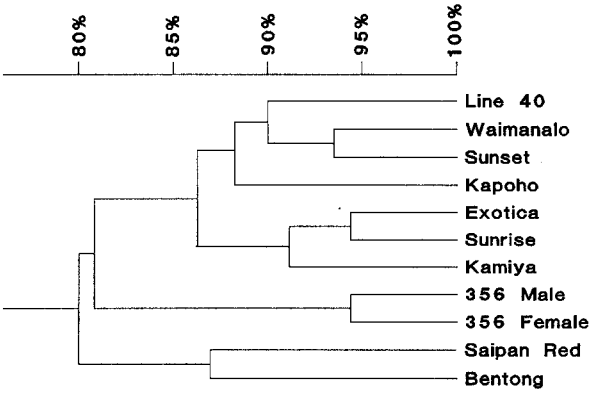

Fig. 2. Dendrogram of papaya cultivars as determined from 102 RAPD probes. The pair-wise coefficients of correlation (Table 2) were clustered using UPGMA and assembled into the dendrogram as described in the methods

Table 2. Simple matching coefficients of similarity determined from analysis using 11 different primers that amplified 102 distinct products. Coefficients are calculated by dividing the sum of the common amplification products present and absent between the cultivars being compared by the total number of fragments observed

	Line-40	Saipan-Red	Kapoho	Waimanalo	Sunset	Exotica	Bentong	Sunrise	Kamiya	356-♂	356-♀
Line-40	1.000										
Saipan-Red	0.776	1.000									
Kapoho	0.864	0.717	1.000								
Waimanalo	0.865	0.773	0.841	1.000							
Sunset	0.900	0.824	0.876	0.920	1.000						
Exotica	0.835	0.858	0.813	0.833	0.905	1.000					
Bentong	0.727	0.863	0.722	0.742	0.793	0.885	1.000				
Sunrise	0.842	0.865	0.781	0.840	0.894	0.947	0.856	1.000			
Kamiya	0.900	0.806	0.815	0.857	0.891	0.905	0.814	0.915	1.000		
356-♂	0.776	0.783	0.772	0.774	0.809	0.788	0.773	0.760	0.755	1.000	
356-♀	0.827	0.777	0.824	0.787	0.840	0.818	0.768	0.700	0.804	0.943	1.000

Jaccard's coefficient of similarity, which excludes common negative data (i.e. where two cultivars each lack the same band), gave a tree in which relationships between cultivars were identical to those determined using simple matching coefficients of similarity. An alternative approach to analysis, parsimony, which groups organisms based on the minimum number of changes needed to explain the observed differences, produced a tree less consistent with the known genetic relationships. The Hawaiian types 'Sunrise' and 'Exotica' were included in a node with 'Saipan Red' and 'Bentong' (data not shown).

Discussion

Our results indicate that RAPD can be used to establish the relationships among closely related cultivars and even individuals in at least some cultivars (Table 2; Fig. 2). We have investigated the use of several different numerical taxonomic techniques for analysis of our data. Since the relationship among seven of our ten cultivars is known, we are able to evaluate various data analysis techniques with regard to their accuracy in constructing relationships among closely related organisms using RAPD data.

Multivariate analysis of RAPD data using simple matching coefficients grouped the seven cultivars having significant Hawaiian components in their genomes and clearly distinguish them from three more distantly related cultivars. Within the Hawaiian group, the deviation of 'Sunset' and 'Kamiya' from expected relationships with 'Sunrise' and 'Waimanalo', respectively, may have resulted from insufficient sampling of the genome since only 11 RAPD primers and only a single plant of each cultivar were used. Confirmation of these data using additional individuals and additional RAPD primers is needed.

The phenogram obtained using parsimony has 'Exotica' and 'Sunrise' on a branch with the cvs. 'Saipan Red' and 'Bentong', and separated from the other Hawaiian cultivars. However, 'Sunrise' is of Hawaiian derivation and 'Exotica' is derived from a 'Subang-6' × 'Sunrise' cross followed by repeated backcrossing to 'Sunrise' to achieve an over 90% expected 'Sunrise' contribution to the genome (Chan 1985; Chan 1987). One might suspect that the small amount of 'Subang-6' germ plasm remaining in 'Exotica' caused this unexpected grouping. The simple matching coefficient method, however, placed 'Exotica' and 'Sunrise' in the same node and within the Hawaiian branch of the tree at a level of similarity about that expected on the basis of the number of backcrosses (Chan 1985; Chan 1987). The unexpected grouping of 'Exotica' and 'Sunrise' with 'Saipan Red' and 'Bentong' by the parsimony method may be an artifact of the method. Parsimony is based on the minimum number of changes to construct a tree. Nodes of the tree are defined by character changes rather than an overall measure of similarity. However, since the plants in this study resulted from, for the most part, specific crosses, an unique amplification product from a minor genomic component may bias the overall tree. Hence, the Malaysian var 'Subang-6' that makes up less than 10% of the 'Exotica' genome may have contributed unique RAPD fragments that placed 'Exotica' with 'Bentong' and biased the placement of 'Sunrise' because of its close relationship with 'Exotica' as a result of repeated backcrossing. The other Hawaiian cultivars group to a common node; however their grouping within the node is slightly different from that given by the simple matching coefficient although it is reasonably consistent with their known genetic relationships.

A second multivariate approach, Jaccard's coefficient, was also used. The phenogram that was produced was identical to that produced using simple matching, except that the similarity coefficients obtained using Jaccard's approach were slightly smaller. Jaccard's approach is related to the simple matching coefficient except that negative data is excluded (Dunn and Everitt 1982; Jaccard 1908). When pair-wise similarity coefficients are calculated, amplification products missing from both of the cultivars being compared are excluded. This correction assumes that, since negative data may be negative for different reasons, it does not indicate a significant correlation. With RAPD data this is probably a valid assumption when the cultivars are not closely related. The greater the genetic distance between the cultivars, the more likely that different genetic changes are responsible for the non-occurrence of a specific amplification product. However, in this study all cultivars have at least a 0.7 simple matching coefficient. Since the cultivars in this study are so closely related, the same sequence variation may be responsible for the non-occurrence of amplification products, making the simple matching coefficient of similarity the most appropriate measure of relatedness.

RAPD analysis revealed only a moderate degree of genetic diversity among the cultivars examined in this study. However, the amount of papaya germ plasm sampled in this study was small and weighted with Hawaiian cultivars. Simple matching coefficients ranging from 0.7 to 0.95 (Table 2) suggest a rather narrow genetic base for domesticated papayas. We are currently in the process of applying RAPD analysis to a broader array of cultivars and wild accessions to get a more complete picture of the genetic diversity within the genus *Carica*.

Acknowledgements. This research was supported by the USDA Special Grants Program for Tropical and Subtropical Agriculture Research.

References

Chan YK (1985) Evaluation of the performance and stability of papaya varieties bred at MARDI. MARDI Res Bull 13:1–7

Chan YK (1987) Backcross method in improvement of papaya (*Carica papaya* L.). Malays Appl Biol 16:95–100

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Dunn G, Everitt BS (1982) An introduction to mathematical taxonomy. Cambridge University Press, Cambridge

Grodzicker T, Williams J, Sharp P, Sambrook J (1975) Physical mapping of temperature-sensitive mutations of adenovirus. Cold Spring Harbor Symp Quant Biol 39:439–446

Hamilton RA, Ito P (1968) Sunrise solo. A different colored solo papaya. Hawaii Agric Exp Stn Circ 69

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaudoise Sci Nat 44:223–270

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Nakasone HY, Crozier JA Jr, Ikehara DK (1972) Evaluation of 'Waimanalo', a new papaya strain. Hawaii Agric Exp Stn Tech Bull 79

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. Nature 335:721–726

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, London

Swofford DL (1990) Phylogenetic analysis using parismony. Version 3.0 Users' manual. Illinois Natural History Survey, Champaign

Welsh J, McClelland M (1991) Genomic fingerprints produced by PCR with consensus tRNA gene primers. Nucleic Acids Res 19:861–866

Welsh J, Honeycutt RJ, McClelland M, Sobral BWS (1991) Parentage determination in maize hybrids using the arbitarily primed polymerase chain reaction (AC-PCR). Theor Appl Genet 82:473

Welsh J, Petersen C, McClelland M (1991) Polymorphisms generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping. Nucleic Acids Res 19:303–306

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acid Res 18:6531–6535