J. Genet. & Breed. 49: 179-186 (1995)

Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of *copia*-like *Eco*RI repetitive elements

rthasarathy

W. Rohde[1]*, **A. Kullaya**[2], **J. Rodriguez**[3] **and E. Ritter**[4]

[1] *MPI für Züchtungsforschung, Carl-von-Linné-Weg 10 D-50829 Köln, FRG,* [2] *National Coconut Development Programme, P.O. Box 6226 Dar es Salaam, Tanzania,* [3] *Albay Research Center, Banao, Guinobatan Albay 4503, Philippines and* [4] *Centro de Investigacion y Mejora Agraria, Apartado 46 01080 Vitoria, Spain*

Received April 7, 1995

ABSTRACT

A novel approach is described for the analysis of coconut germplasm by the use of coconut-specific primers complementary to the previously identified highly repetitive *copia*-like *Eco*RI elements. PCR amplification of spacer regions for a subset of tandemly arranged repeats detected polymorphisms which allowed an analysis of biodiversity within coconut populations. In addition, genetic relationships were determined for selected coconut types from various regions of the world. The results support the theory on human-assisted dissemination of coconut germplasm.

Key words: Biodiversity; Coconut; Genome analysis; Human-assisted dissemination; PCR.

INTRODUCTION

The coconut palm (*Cocos nucifera* L.) is one of the major perennial oil crops of the tropics providing the basis in many developing countries for food products as well as serving industrial purposes (Jones, 1991; Persley, 1992). It is basically a smallholder's crop with some 96% of all coconuts being produced by small farmers owning less than 4 ha of land, and, therefore, it is an important attribute of the rural economy. In addition, coconut has a strong economical part to play such as in the export of coconut oil or copra in the case of the Philippines or for national consumption as for example in Indonesia which over the past three years has become the world's largest producer of coconut oil, most of which is consumed in the country. However, owing to the relatively low productivity of the coconut palm, the increased demand for timber, a strong competition by other perennial oil crops like the oil palm (especially by its byproduct palm kernel oil) as well as due to devastating diseases caused by various pathogens like viroids (cadang-cadang disease), viruses (foliar decay disease) or mycoplasma-like organisms (lethal disease, lethal yellowing), coconut cultivation is in decline in many countries.

As a consequence, breeding efforts in coconut have to be directed towards increasing the yield by breeding early-flowering, high-producing hybrids with resistance or tolerance to biotic and abiotic stresses. Secondly, planting or breeding material has to be certified by indexing for pathogens before release to the farmers or for the safe movement of germplasm (Frison *et al.*, 1993). For breeding purposes the selection of genetic material in crosses has to accomodate for the genetic variability encountered in germplasm collections. With the advent of DNA marker technology, the characterization of genetic diversity in coconut germplasm at the DNA level has only recently begun (Ashburner and Rohde, 1994; Rohde, 1995) to substitute other strategies like isozyme or leaf polyphenol analysis. Both the analysis of genetic variability among parents used in breeding programmes and pathogen diagnosis can these days rely on highly sensitive, non-radioactive methods for the identification of plant or pathogen genomes (Rafalski and Tingey, 1993). These include the rapid amplification by the polymerase chain reaction (PCR) of DNA or RNA sequences with the latter requiring a first step of reverse transcription (RT-PCR) in the presence of specific primers, or by using random primers in the random amplification of polymor-

---

* Corresponding author.

phic DNA (RAPD). Furthermore, variation in DNA sequences can be documented by digestion of DNA with restriction endonucleases followed by Southern analysis with DNA markers which detect restriction fragment length polymorphism (RFLP).

One additional strategy is based on the presence of mini- or microsatellite DNAs in plant and animal genomes. These satellite DNAs represent highly repetitive, hypervariable regions of varying length which were first discovered in the human genome (Jeffreys *et al.*, 1985) and are flanked by conserved sequences (sequence-tagged microsatellites, STMs). Primers complementary to these conserved sequences allow the amplification of STMs by PCR and result in a DNA fragment pattern specific for a particular genotype. For coconut we have previously described a family of highly repetitive sequences which are generated by *Eco*RI restriction of genomic DNA and show a remarkable sequence homology to the *copia*-like group of retroposons (Rohde *et al.*, 1992). The use of these repetitive sequences as potential RFLP markers in molecular hybridization experiments, however, failed to reveal polymorphism among a variety of coconut types (Rohde, 1995). Here we describe a novel strategy to detect polymorphism by inverse sequence-tagged repeats (ISTR) analysis in that primers complementary to the repetitive sequences were used for PCR to amplify spacer regions between individual *Eco*RI repetitive units. The implications of the results on the analysis of biodiversity in coconut germplasm and for breeding strategies are being discussed.

MATERIALS AND METHODS

Plant material

The coconut accessions used in this study are grown at various locations in Tanzania and the Philippines under the auspices of the National Coconut Development Programme (NCDP), Tanzania, and the Philippine Coconut Authority (PCA), Philippines. Material was harvested from the emerging leaf and kept without freezing until DNA extraction.

Isolation of genomic DNA

The isolation of genomic DNA was performed by a modification of the procedure described by Doyle and Doyle (1990). 10 g of leaf material (midribs removed) were cut into small pieces and ground with mortar and pestle in 50 ml of CTAB extraction buffer (100 mM TRIS/HCl pH 8, 1.4 M NaCl, 20 mM EDTA, 2% CTAB + 0.2% freshly added mercaptoethanol) for 3-5 min in the presence of 500 mg of acid-washed sea sand. The suspension was transferred to a small glass bottle and incubated at 60 °C for 30 min with occasional stirring. The incubation mixture was filtered through Miracloth with thorough squeezing, the filtrate divided into two 50ml-Falcon tubes, extracted with an equal volume of chloroform and the phases were separated by centrifugation at 4000 rpm for 5 min.

Nucleic acids were precipitated from the clear supernatant by the addition of 0.8 volumes of isopropanol at room temperature for 5 min and then collected by centrifugation at 4000 rpm for 10 min. The supernatant was discarded, and the pellet was dried briefly in the air and then dissolved in 5 ml of TE buffer (10 mM TRIS/HCl, pH 8.0, 0.1 mM EDTA) containing 20 μg/ml RNase A. After incubation for 60 min at 37 °C the aqueous phase was extracted with an equal volume of phenol followed by a chloroform extraction, and the DNA was precipitated after the addition of 1/10 volume of 3 M NaOAc with 2.5 volumes of ethanol or 0.8 volume of isopropanol. The DNA was collected by centrifugation at 4000 rpm/10 min, the pellet was washed with alcohol, air-dried and redissolved in 1 ml of TE buffer.

The concentration of DNA was determined spectrophotometrically and its yield was generally in the range of 0.5 mg/10 g of leaf tissue. The integrity of the DNA was checked on a 0.7% agarose gel, and it consisted of high molecular weight DNA without visible degradation. It was readily digested by restriction enzymes and, therefore, directly suitable for RFLP analyses. For PCR amplification, however, an additional purification step by chromatography on QIAGEN colums (QIAGEN, Hilden, FRG) or comparable reversible adsorbents was necessary in order to obtain reproducible results.

Amplification of coconut genomic DNA

Primers were designed that were complementary to the 5' ends of the two *Eco*RI repetitive sequences Ecorep1 (#2907) and Ecorep3 (#2892) and to a sequence within the 3' end common to both elements (#2903; Fig. 1A). PCR reactions were performed according to standard protocols in a final volume of 25 μl containing 25 ng of genomic DNA, 200 μM dNTPs, 2.5 mM $MgCl_2$, 1 x PCR buffer (Gibco/BRL), 20 nM primers (2907/2903 or 2892/2903 primer pair), and 1 unit of Taq DNA polymerase (Gibco/BRL). The programme consisted of the following steps: step 1, 95 °C/3 min; step 2,95 °C/30 sec; step 3,45 °C/30 sec; step 4,72 °C/2 min; step 5,72 °C/10 min with 40 cycles of steps 2 to 4.

Processing of PCR products

For sequence analysis the PCR products were isolated after agarose gel separation by reversible adsorption to QIAEX, phosphorylated with polynucleotide kinase and subcloned into the *Sma*I site of Bluescript IIKS. Recombinant clones were sequenced by cycle sequencing using either ^{32}P-labelled primers or fluorescence-labelled chain terminators in a semiautomatic sequencer.

For ISTR analysis, the primers were labelled with polynucleotide kinase and [γ-^{33}P] ATP before PCR amplification under the conditions described above. Reaction products were separated on a 4% polyacrylamide gel and made visible by autoradiography.

Computer analysis

Polymorphic fragments visible in the ISTR gels were scored for their presence (1) or absence (0) in the respective genotypes. The data were processed using the NTSYS-pc programme (Exeter Software, Setauket, USA). Simple matching coefficients were calculated between each pair of genotypes using the quotient numer of matches/total sample size. Based on the matrix of these coefficients between all genotypes, the cluster analysis was performed by the UPGMA method.

RESULTS

Arrangement of repetitive elements in the coconut genome

In a first series of experiments, the coconut genome was analysed for a possible tandem arrangement of the previously identified *copia*-like *Eco*RI repetitive elements (Rohde *et al.*, 1992). For this, three primers complementary to the elements as shown for Ecorep1 in Fig. 1A were employed in PCR amplifications of genomic DNA from the three individual palm types East African Tall (EAT), Pemba Red Dwarf (PRD) and Rennell Tall (RLT). These trees were chosen as they are parents in ongoing breeding programmes for the establishment of a coconut RFLP and RAPD map. Only primer pair 2907/2903 yielded a PCR product of approximately 400 bp (data not shown) which was isolated for all three coconut types and subcloned. Sequences were established for several clones of each coconut genotype.

A pileup of these sequences is shown in Fig. 1B and reveals several interesting features. Firstly, the amplified regions can be classified into two groups differering by the occurrence of two internal deletions of 9 and 15 bp, respectively, in one group, while the other is shorter by 21 bp at the 3' end and contains an internal deletion of 3 bp resulting in an overall identical length for most of the spacers. However, other deletions occur in both groups and one spacer carries a deletion of 34 bp at its 3' end. A second remarkable feature is the absence of the left *Eco*RI site that was expected to be present around position 380 (Fig. 1B) on the amplified spacers because of the location of the primers (Fig. 1A), while the right *Eco*RI site is present in most, but not all spacers (Fig. 1B). This finding excluded the possiblity for a tandem arrangement of the previously identified highly repetitive *Eco*RI elements in coconut and suggested that a subset of these repeats exists which display varying lengths and are arranged in tandem. This result corroborates our previous finding on the existence in the coconut genome of minor bands of 2.5, 2.7, 2.9 and 4 kb which hybridize to the Ecorep sequences (Rohde *et al.*, 1992).

Analysis of spacer regions as a measure for genetic diversity in coconut populations

On the basis of the results obtained above, it appeared possible to utilize the existence of deletions in the amplified regions for analyzing coconut germplasm with respect to biodiversity and to assess the genetic homogeneity of populations of the identical coconut type. Therefore, the primers were radiolabelled before PCR amplification of coconut DNAs and the products obtained with both primer pairs were analyzed on a 4% polyacrylamide gel. The EAT, PRD and RLT palms used above in PCR amplification plus six additional EAT palms from the identical Tanzanian EAT population were subjected to this first analysis. Amplification with both primer pairs resulted in the PCR patterns shown in Fig. 2A. It was evident that the EAT population was highly heterogeneous as seen by the absence or presence of single bands in the PCR patterns generated for the seven individual EAT palm trees. Also, the single PRD and RLT trees (the pollen donors in the respective crosses with the seven EAT mother palms) gave rise to unique PCR patterns which allowed to distinguish between all parents used in the crosses.

Dwarf coconut types are assumed to be autogamous plants, as the receptivity of the female flower coincides with pollen maturation on the same inflorescence. Therefore, cross-pollination is thought to be largely suppressed. In order to test this hypothesis, ISTR analysis was extended to 10 and 17 individual trees from two Malayan Yellow Dwarf (MYD) populations grown at Selem (Zanzibar) and Albay (Philippines), respectively. The results (Fig. 2B) can be summarized as follows. Although most of the PCR products were of identical size, the MYD populations were not identical, and, secondly, even within a given population genetic diversity was evident as most clearly detected for the MYD palm trees from the Philippines.

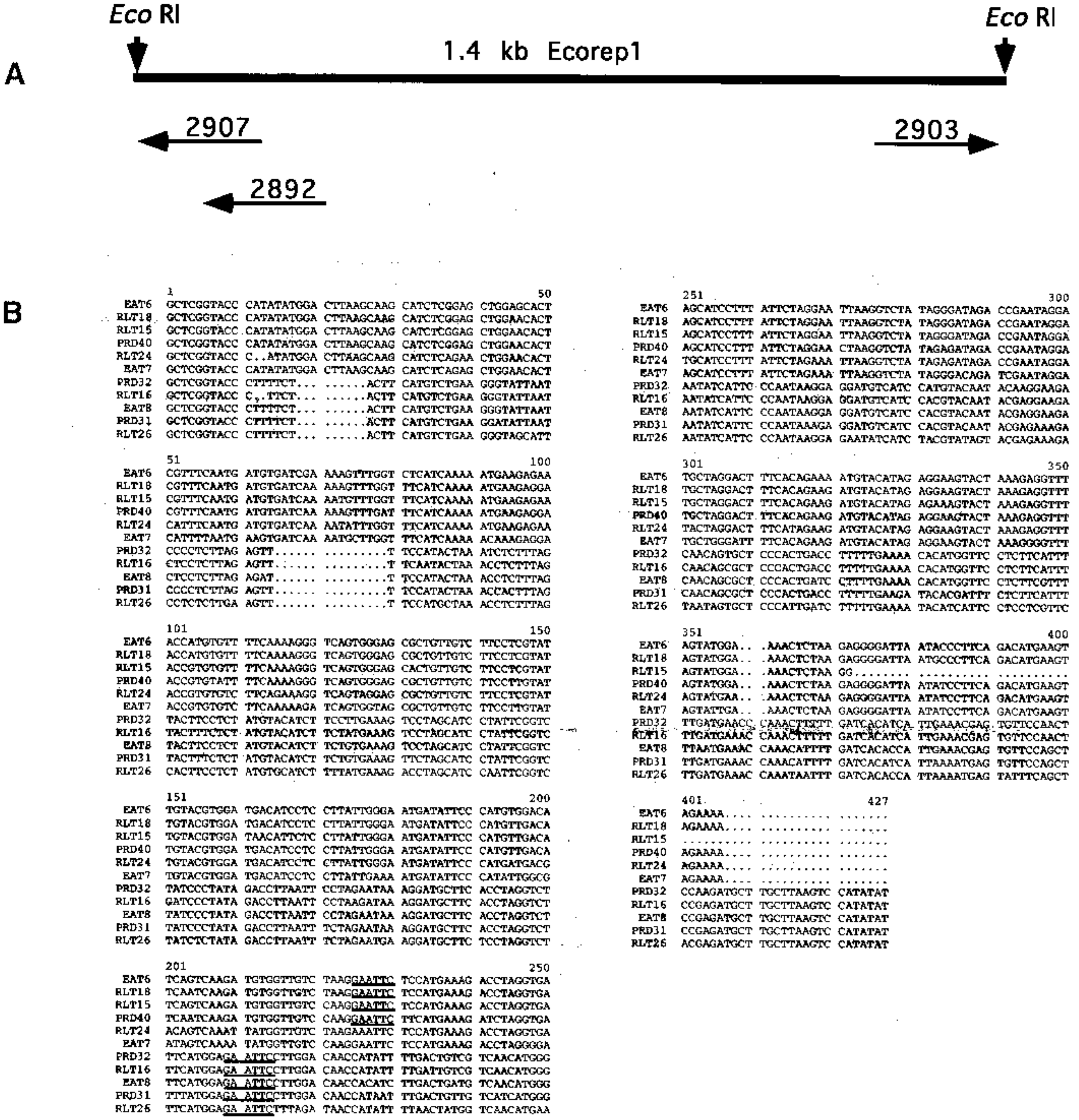

FIGURE 1 - Sequence analysis of spacer regions separating a subset of *Eco*RI repeat units. (A) Location of the primers used for PCR amplification on the 1.4 kb Ecorep1 sequence. The terminal *Eco*RI sites are indicated. (B) Sequence pileup of regions amplified for the three coconut types East African Tall (EAT), Pemba Red Dwarf (PRD) and Rennell Tall (RLT). The numbers associated with each coconut type indicate individual clones. Internal *Eco*RI sites corresponding to the 3' located *Eco*RI site of Ecorep1 (A) are underlined.

ISTR as a tool to study genetic relationships among different coconut types

The analysis of the coconut genome was applied to 21 coconut types that were available in the Tanzanian germplasm collection and had been accessed from different regions of the world. These included Latinamerican, African, Asian and Polynesian material which was subjected to ISTR PCR amplification with the 2892/2903 primer pair. A total of 36 polymorphic bands were distinguished without ambiguity (data not shown) from which a

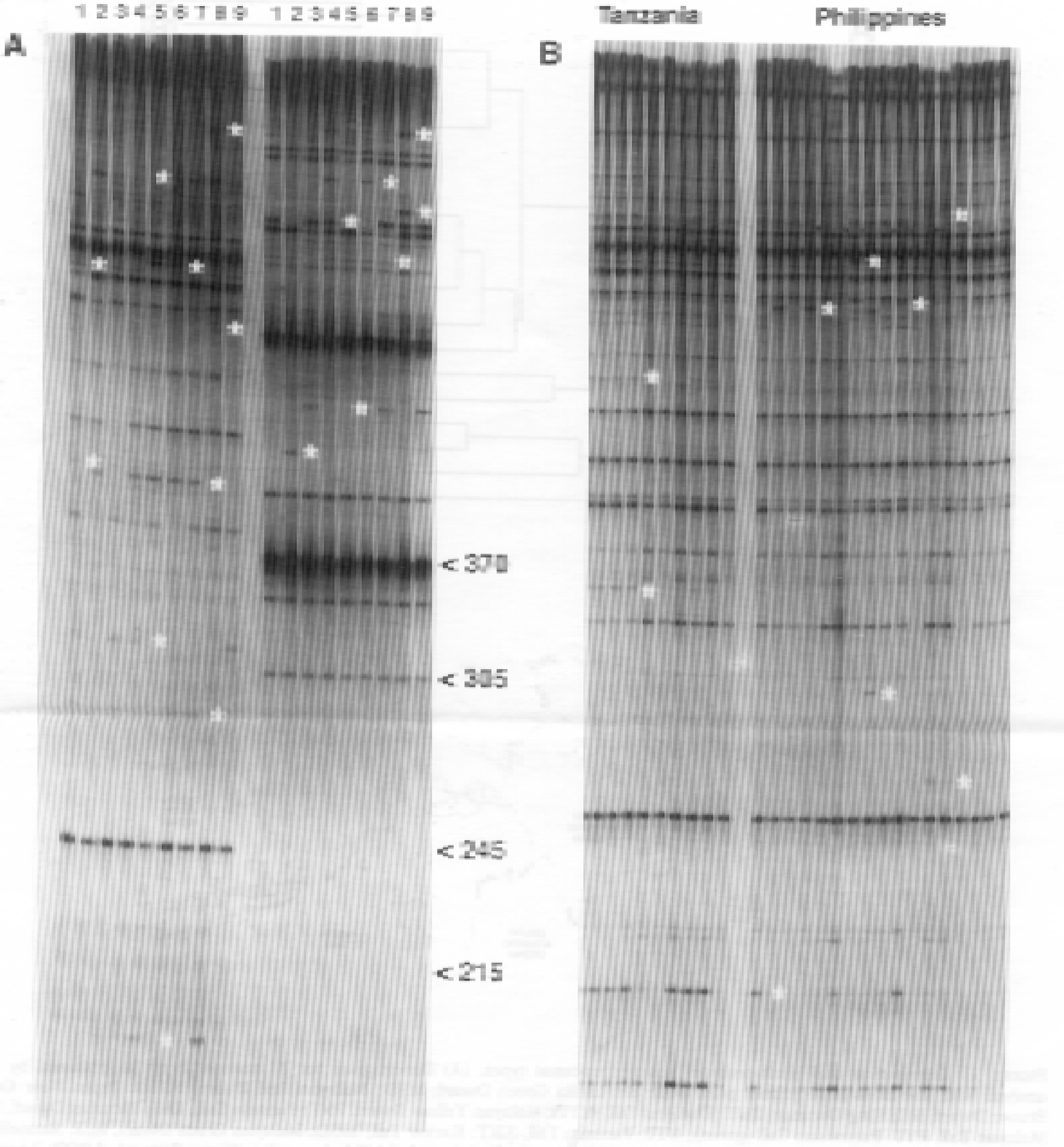

FIGURE 2 - ISTR analysis of coconut populations. (A) Analysis of a Tanzanian EAT population (numbers 1 to 7) as well as single RLT (#8) and PRD (#9) palm trees with the two primer pairs 2892/2903 (left panel) and 2907/2903 (right panel). ^{33}P-labelled primers were used in a PCR amplification, the products were separated on a 4% polyacrylamide gel and made visible by autoradiography as described in the Materials and Methods section. (B) ISTR analysis of Malayan Yellow Dwarf (MYD) populations from Tanzania and the Philippines. The primer pair used was 2892/2903. Asterisks denote most prominent polymorphic PCR products. The indicated size markers (in nucleotides) were established from a sequence reaction run on the same gel.

matrix was constructed and a dendrogram obtained (Fig. 3A). This dendrogram has to be regarded as provisional, as only a single tree per accession was examined (high genetic variability among tall coconut types; see above and Fig. 2A) and the number of polymorphic markers detected by the primer pair was relatively low. Nevertheless, some interesting genetic relationships became evident. Coconut palms from Africa (group I) have clearly separated early during evolution from the

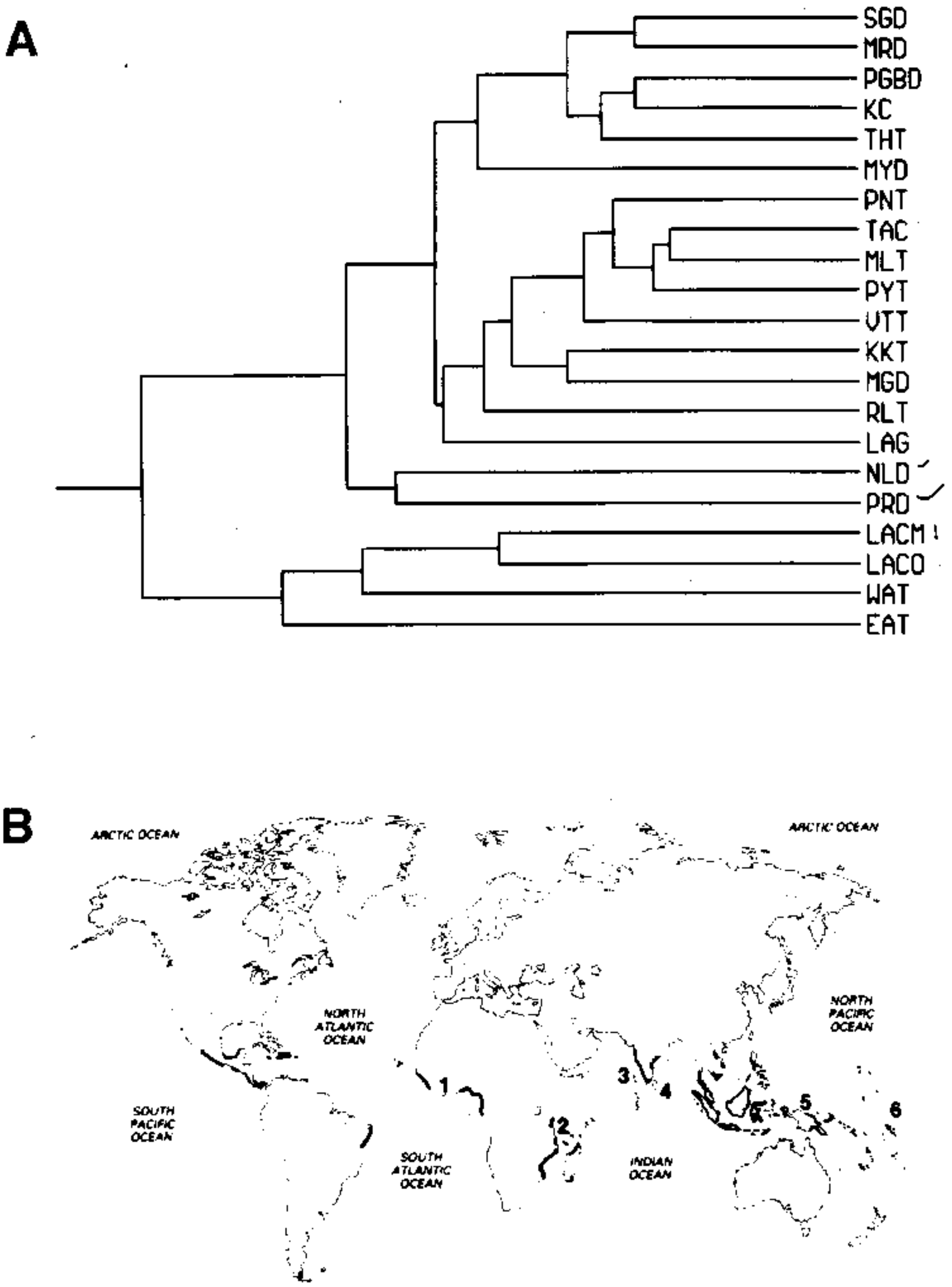

FIGURE 3 - Distribution and biodiversity of selected coconut types. (A) Dendrogram for 21 coconut types as obtained by ISTR analysis with the 2892/2903 primer pair. SGD: Sri Lanka Green Dwarf, MRD: Malayan Red Dwarf, PGBD: Papua New Guinea Brown Dwarf, KC: King Coconut, THT: Thailand Tall, MYD: Malayan Yellow Dwarf, PNT: Panama Tall, TAC: Tacunan Dwarf, MLT: Malayan Tall, PYT: Polynesian Tall (green), VTT: Vanuatu Tall, KKT: Karkar Tall, MGD: Malayan Green Dwarf, RLT: Rennell Tall, LAG: Laguna Tall, NLD: Niu Leka Green Dwarf, PRD: Pemba Red Dwarf, LACM: Laccadive Micro (brown), LACO: Laccadive Ordinary (brown), WAT: West African Tall, EAT: East African Tall. (B) Distribution of several selected coconut types with genetic relationships. 1: WAT, 2: PRD, 3: LACM, LACO, 4: KC, 5: PGBD, 6: NLD. The map was taken from PLUCKNETT (1979) and modified. Darkened lines indicate coconut-growing regions.

other analyzed accessions (group II). There are, however, several exceptions as highlighted in Fig. 3B. The two ecotypes LACM and LACO (see legend to Fig. 3 for abbreviations) collected from the Laccadive Islands off the Southwest Indian coast are highly related to the African coconut types EAT and WAT. On the other hand, the PRD palm collected on Pemba Island off the Tanzanian coast is not an African coconut type, but belongs to the Polynesian coconut germplasm due to its high homology with NLD from Fiji. Another exception is remarkable and occurs in a group II subgroup comprising Thailand, Malaysia and Sri Lanka: The PGDB palm from Papua New Guinea is closely related to King Coconut (KC) from Sri Lanka.

DISCUSSION

The results described here bear relevance to coconut breeding programmes, as they allow to determine genetic diversity among coconut types assembled in germplasm collections, and thus help in the selection of parents for hybrid production. In the specific novel experimental approach which is described here and which we have termed ISTR, we make use of the previously identified, highly repetitive *Eco*RI family of *copia*-like retroelements in the coconut genome (Rohde *et al.*, 1992). It was established that a subset of elements exists which are arranged in tandem with the occurrence of deletions in the spacer as well as in the element regions amplified by selected primers in a PCR reaction. In contrast to random primers in RAPD analyses, the ISTR primers are coconut-specific and will, therefore, not amplify bacterial or fungal contaminations on leaf material. Secondly, the DNA polymorphisms can be increased by the use of additional primers complementary to different regions of the spacers and/or repeat units.

This analytical approach was applied to an EAT population used in an ongoing breeding programme at NCDP, and it confirmed the expected existence of genetic diversity among the tall coconut type. More importantly, however, in light of the generally accepted view of dwarf coconut types as being autogamous, two MYD populations growing in Tanzania and the Philippines show diversity between as well as within populations. The MYD coconut type has been a favourite parent in hybrid production, and the Maypan hybrid obtained in Jamaica from a MYD x PNT (Panama Tall) cross has been reported to show resistance to mycoplasma-like organisms (MLO), while in other locations hybrids from comparable crosses with MYD failed to suppress MLO-caused lethal yellowing or lethal disease (Burotrop, 1992). It cannot be excluded that in Jamaica a different MLO strain is prevalent, but our results open the possibility for genetically determined MLO resistance inherited by the particular MYD parents that were used in the Jamaican crosses. As the MLO-caused diseases (lethal yellowing, lethal disease) are considered to pose the most devastating threat to coconut production (Persley, 1992), the search for DNA markers cosegregating with MLO resistance will be of utmost importance.

The analysis of a small number of coconut accessions from different parts of the world allowed to draw interesting conclusions on the validity of theories on the natural and human-assisted spread of coconut germplasm (Purseglove, 1972; Harries, 1978). The fact that ecotypes from the Laccadive Islands, which are located off the Southwest coast of India, are related to the coconut palms from Africa would suggest that this occurred by floating with favourable water currents or that in the course of trading coconut germplasm was spread from the East African coast to India (Fig. 3B). This latter mode of dissemination may have happened either by Arabic merchants who were originally dominating the exchange of goods between East Africa and Calicut on the Southwest Indian coast (Purseglove, 1972). Alternatively, as a concequence of Vasco da Gama's opening in 1499 of the route to India by sailing around the African coast, sea-faring nations like the Portuguese or Spaniards may have disseminated coconut germplasm to the Laccadives, an alternative also discussed by Harries (1977). This explanation appears likely in view of the fact that the Portuguese have introduced coconut from East Africa into West Africa at the end of the fifteenth century (Lemothe and Sangare, 1993). The spread by merchants making use of coconuts for refreshment during their voyage may also explain why the PGBD palm from Papua New Guinea is highly related to the King Coconut from Sri Lanka which is found in the South-East Asian subgroup of coconut germplasm, and similar conclusions apply to the clustering of the Malaysian coconut types MLT and MGD or the Panama Tall (PNT) in the Polynesian subgroup. In addition, the high genetic relationship between PRD from Pemba Island (Tanzania) and the Polynesian germplasm can only be explained by distribution through man and not by natural floating. In any event, these first data on biodiversity among different coconut accessions by DNA marker technology presented here, lend support to the importance of human-assisted dissemination of coconut germplasm (Harries, 1978).

One additional aspect of these studies deserves attention. Except for coconut foliar decay virus (CFDV) for which appropriate experiments have not yet been completed, the nucleic acid extraction procedure from coconut leaves described here allows the simultaneous analysis of coconut genomic DNA as well as its most important pathogens like MLO and cadang-cadang viroid (CCCVd) or viroid-related sequences provided that the RNAse step is omitted from the protocol (unpublished). Furthermore, the mechanical destruction of leaf material can be performed with a hammer on leaf material plus buffer placed in polyethylene bags and followed by the 60 °C incubation step, chloroform extraction and isopropanol precipitation of nucleic acids. This procedure is applicable everywhere with a minimal amount of equipment and funds. It is, therefore, foreseen than regional centers like NCDP or PCA ARC will be in the position to assist neighbouring countries in their breeding and disease control programmes by the analysis of coconut germplasm and the indexing for pathogens without the restriction by current quarantine regulations (Frison *et al.*, 1993).

ACKNOWLEDGEMENTS

The authors are grateful to Dieter Becker and Alice Kaufmann for their technical assistance and to Dr. Ralph Schäfer-Pregl for help with the computer analysis. This work was funded under contract No. TS3*-CT93-0213 of the EC STD3 programme.

REFERENCES

Ashburner, G.R. and Rohde, W., 1994. Coconut germplasm characterisation using DNA marker technology. *ACIAR Proceedings*, **43**: 44-46.

Burotrop, 1992. *Research and Development of Coconut in Latin America and the Carribean.* Seminar Proceedings, Kingston, Jamaica.

Doyle, J.J. and Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. *Focus*, **12**: 13-15.

Frison, E.A., Putter, C.A.J. and Diekmann, M., eds., 1993 *FAO/IBPGR Technical Guidelines for the Safe Movement of Coconut Germplasm* FAO/IBPGR, Rome.

Harries, H.C., 1977. The Cap Verde region (1499 to 1549); key to coconut culture in the Western hemisphere? *Turrialba*, **27**: 227-231.

Harries, H.C., 1978. The evolution, dissemination and classification of *Cocos nucifera* L. *Bot. Review*, **44**: 265-317.

Jeffreys, A.J., Wilson, V. and Thein, S.L., 1985. Hypervariable "minisatellite" regions in human DNA. *Nature*, **314**: 67-73.

Jones, L.H., 1991. Perennial vegetable oil crops. In: Persley, G.J., ed., *Agricultural Biotechnology: Opportunities for International Development.* CAB International, Wallingford, UK, pp. 213-224.

Lamothe de Nuce De, M. and Sangare, A., 1993. Current status of coconut genetic resources research in the Ivory Coast. In: Nair, M.K., Khan, H.H., Gopalasundaram, P. and Bhaskara Rao, E.V.V., eds., *Advances in Coconut Research and Development.* Oxford & IBH Publishing Co., New Delhi, India, pp. 9-14.

Persley, G.J., 1992. *Replanting the Tree of Life. Towards an International Agenda for Coconut Palm Research.* CAB International, Wallingford, UK.

Plucknett, D.L., 1979. *Managing Pastures and Cattle under coconuts.* Westview Press, Boulder, USA.

Purseglove, J.W., 1972. *Tropical Crops. Monocotyledons.* Longman, London, UK.

Rafalski, J.A. and Tingey, S.V., 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.*, **9**: 275-280.

Rohde, W., 1995. Improvement of coconut by biotechnology: application of molecular techniques to breeding and crop protection. In: "*La Recherche Européene au Service du Cocotier - Actes du Séminaire - 8-10 septembre 1993, Montpellier*". CIRAD (Collection: Colloques du CIRAD), Montpellier (in press).

Rohde, W., Salamini, F., Ashburner, R. and Randles, J.W., 1992. An *Eco*RI repetitive sequence family of the coconut palm *Cocos nucifera* L. shows sequence homology to *copia*-like elements. *J. Genet. & Breed.*, **46**: 391-394.