Origin, Domestication, Dissemination and Genetic Diversity of Coconut:
DNA information

By
Lalith Perera

Abstract

Information on origin, dissemination and levels and distribution of genetic diversity in coconuts will allow plant breeders and conservationists to select better breeding materials and formulate appropriate conservation strategies. PCR-based DNA profiling of coconut palms from Sri Lanka was initially conducted using both Amplified Fragment Length Polymorphism (AFLPs) and Microsatellites (SSRs). Thirty-nine microsatellite primers specific to coconut were developed by small insert genomic library construction. Eighteen of those primers were used to analyze the same set of Sri Lankan coconut materials. Overall, the results generated by both AFLPs and SSRs were in agreement. Most diversity was found in the tall variety (Typica) (0.92 and 0.62 for AFLPs and SSRs, respectively) rather than the intermediate (Aurantiaca) and dwarf (Nana) varieties (0.82 and 0.25 for AFLPs and SSRs, respectively). A hierarchical analysis of molecular variance (AMOVA) based on AFLP data was used to quantify and partition levels of variability between and within form components. This revealed that for the inbreeding dwarf and intermediate forms most variation was observed between rather than within forms. In contrast, the out-breeding tall forms exhibited as much variation within as between forms.

Under the coconut bio-diversity conservation programme of Sri Lanka, several collections of tall coconuts were planted ex-situ. Thirty-three populations were subjected to SSR assay with eight microsatellite primers. A high level of genetic diversity (0.68) and high level of within population variation (98%) were detected. There was very little population differentiation. It was concluded that a single large collection would adequately represent the genetic diversity in the targeted area. These results generated useful information for the coconut bio-diversity conservation programme in Sri Lanka.

Twelve pairs of SSR primers were used to screen worldwide collections of coconut germplasm. Eighty-four alleles were detected in tallss compared with only 42 in dwarfs and average diversity value in tallss was significantly higher (0.703) than in dwarfs (0.374). It was concluded that dwarfs are a subset of the tall coconuts and are directly evolved from tall and moreover from tallss from Southeast Asia and Pacific. These results provided evidence in support of previous hypotheses concerning the dissemination of coconut, as well as important new information for conservation and breeding purposes. Levels of cytoplasmic diversity in coconut were assessed by using both restriction digestions of PCR products in the chloroplast and mitochondrial genomes and chloroplast SSRs. No variation was found suggesting coconut may have gone through a severe cytoplasmic bottleneck.

The information emerging from this study enhances the knowledge of amount and distribution of genetic diversity, origin, evolution and dissemination of coconut. The use of the results generated for formulating conservation strategies, selection of parents for the breeding programmes as well as strategic planning of genome mapping project of coconut are discussed.

The coconut palm, Cocos nucifera L, was a major plantation oil crop in the world until the mid 1960s but it continues to be important in developing countries where there is a high demand for coconut oil in the domestic market. In addition to its economic importance, coconut plays a significant role in the cultural and social life of people in over 80 countries. Because of these factors

* Genetics and Plant Breeding Division, Coconut Research Institute, Lunuwila, Sri Lanka
coconut will continue to be important where it is grown. But the fact that coconut oil contains a high percentage of industrially valuable lauric acid means that it has to compete with palm kernel oil from oil palm as well as oils from genetically modified crops such as canola. Therefore, a continued effort to improve coconut is required if it is not to be totally displaced from world markets. Genetic improvements in coconut have been confined to mass selection in the field and inter and intra-variety hybridization. Evaluation and selection of coconut planting material is currently based on morphological traits; and, although these do not necessarily reflect variation in coconut genome, significant advances in improvement of coconut palm have been achieved through conventional breeding. Yet, the scope for improvement is large and is to be achieved by combining traditional plant breeding methods with modern biotechnology. Therefore, the immediate requirement to improve coconut palm is to incorporate molecular tools into the breeding programme. DNA marker technology allows the accurate measurement of genetic diversity and its distribution, independent of the environment. This, in turn, enables the construction of genetic linkage maps to identify marker trait associations for use in marker-assisted selection and breeding.

Currently, there are many powerful molecular marker systems. The choice of markers depends upon several factors such as the information content of the marker, ease of performance, reproducibility, expense and available skills and facilities. Among marker systems, the most promising are AFLPs (high throughput) and SSRs (highly polymorphic) \( (\text{Powell} \; et \; al., \; 1996ab) \). The AFLPs can generate large quantities of information (high multiflex ratio) in a relatively short time and do not require prior sequence information on the crop. However, AFLPs are dominant markers and, therefore, relatively less informative than co-dominant markers. SSRs are the most informative marker systems today. They are multi-allelic, co-dominant, highly polymorphic and abundant. They are, however, developmentally expensive because prior sequence knowledge of the crop concerned is essential.

Many species, particularly perennial tree species such as coconut, are extremely difficult to collect and maintain \textit{ex-situ}. Therefore, it is extremely important to measure the amount and distribution of genetic diversity prior to deciding on where to collect and how many samples to collect. In addition, evaluation of new coconut hybrids to confirm their performance may take 15 or 20 years; and, therefore, the right choice of materials to cross in the breeding program, in order to maximize heterosis, is important. At present, there is very little information on the levels of genetic diversity and its distribution between varieties and between geographical regions. The genetic behaviour of characters and their interaction with one another are not known. The long generation interval in coconut (7-10 years) means that assessment of the most desirable quantitative traits such as precocity and yield (time to flower, nut number, nut size, copra and oil) is a lengthy process (15-20 years). Development of inbred lines, through selection cycles, is virtually impossible as a minimum of six selfing cycles would take around 60 years to complete. In addition, coconut requires large areas of land for evaluation. Screening large numbers of palms in a single field experiment is a difficult and a costly task. The lack of knowledge of the amount of genetic diversity present in a population, because it is not accurately reflected by the morphological diversity and quantitative variation, will slow down the improvement that can be achieved through selection. Similarly, a lack of knowledge of the true genetic relationships between palms restricts selection of better parental combinations for increased heterosis. In contrast, genetic markers provide accurate estimates of genetic relationships between palms. Moreover, identification of markers linked to phenotypic traits means that breeders can begin to manipulate genes, even without knowing the exact location of the gene. Furthermore, desirable genotypes can be selected at an early vegetative stage long before the characters are expressed (7-10 years for reproductive characters in coconut) and recessive alleles can be identified even when they are not expressed in the phenotype. These applications to the coconut palm will accelerate the improvement of this crop and shorten its breeding cycle.

DNA profiling of coconut palms is relatively recent. The application of molecular markers in coconut was first reported in 1996 with the use of RAPDs to evaluate Sri Lankan coconut population (Everard, 1996) and South Pacific coconut populations (Ashburner \textit{et al.}, 1997). The first publication
on the application in coconut of AFLPs (Perera et al., 1998) and SSRs (Perera et al., 1999), respectively, was by the author of this article. Since no published SSR primers were available for coconut, a small insert genomic library was constructed and enriched for (CA)_n repeats using the procedure described by White and Powell (1997). Thirty-nine SSR primers were developed for coconut and were shown to be multi-allelic (2 for CAC21 to 11 for CAC56) and highly polymorphic (gene diversity 0.39±0.04 for CAC3 to 0.81±0.01 for CAC56). Eight AFLP primer combinations and 18 SSR primers were used to evaluate all the varieties of coconut in Sri Lanka. A high level of genetic diversity was observed for Sri Lankan coconut with overall genetic diversity exceeding 0.9 for AFLPs and 0.6 for SSRs. The results of AFLPs were largely in agreement with the results obtained for SSRs. In both studies, most variation was detected in the predominantly out-breeding tall (Typica) coconuts rather than the intermediate (Aurantiaca) and dwarf (Nana) coconuts that are predominantly inbreeding (Liyanage, 1958). Most variation was observed between, rather than within, dwarf and intermediate populations. In contrast, the tall coconut types exhibited as much variation within populations as between. These observations have important implications for the collection and maintenance of coconut germplasm resources, revealing that, for tall coconuts, emphasis should be given to both number of individuals and the number of populations; whereas, for dwarfs and intermediates, the number of population is the important factor.

The analysis of thirty-three Sri Lankan coconut populations belonging to the commercially grown variety of coconuts in Sri Lanka using eight SSRs revealed a very high level of within population variation (98.5%). This indicates that there are no distinctive populations within the native range of coconuts in Sri Lanka. High levels of within population variation is a common observation in other out-breeding crops also, for example in *Theobroma cacao* (Allen, 1988; Russell et al., 1993) and tropical tree species *Calycophyllum spruceanum* (Russell et al., 1999) where more than 90% of the variation was observed within population. It is likely that there is a common history and a narrow genetic base for commercially grown Sri Lankan tall coconuts. These results stress the importance of prior knowledge on the amount and distribution of genetic diversity for appropriate collection and conservation strategies. Previous collections have been based on morphology and quantitative data and the assumption that populations grown in different environments are subject to different selection pressure and are adapted differently. The criterion for germplasm collection based on the *in-situ* measurement (growth habit and the fruit component data) alone is not a particularly useful measurement of genetic diversity as the environment largely influences them. Therefore, combining phenotypic data particularly fruit component data in coconut with molecular data would be the appropriate strategy. The molecular data generated in this study and other such studies will help formulate accurate collection strategies. Individual strategies must be considered for different regions based on the level and distribution genetic diversity in the particular area evaluated using molecular markers. In the case of commercially grown Sri Lankan tall coconut, it can be concluded that a single large random collection could serve as a representative of all these coconuts. Moreover, the materials already collected and conserved *ex-situ* adequately represents total genetic variation. In contrast, it is suggested that collection of many populations would be the appropriate strategy for the South-Pacific region, where approximately 40% of the variation, based on RAPD data has been found between populations (Ashburner et al., 1997). Future studies involving molecular sampling of all the coconut growing areas is necessary to devise collection strategies; and, to this end, collaboration between research groups are desirable.

Molecular markers also provided new insights into genetic relatedness as compared to phenotypic relatedness of coconut forms. For example, morphologically, the aurantiaca group of coconut in Sri Lanka is considered to be intermediate between the tall and dwarf varieties but the estimation of genetic relatedness, based on AFLP and SSR analysis, identified the aurantiaca group as being closer to the dwarf group than to the tall group. The variety bodiri of Sri Lanka, classified as a tall variety (Liyanage, 1958) based on morphology was found to be more similar to dwarf types than tall types, based on their SSRs relationships. Similarly, dwarfs were found to be closer to the tall coconuts from Southeast Asia and Pacific than to the tall coconuts of South Asia. In addition, putative duplicate genotypes were identified in the aurantiaca group of coconuts. Identification of
duplicate accessions is an important issue in germplasm conservation as maintenance of duplicate accessions is costly, especially when living specimens are maintained as a field planting.

Information emerging from this SSR and AFLP study facilitates the management of coconut germplasm and allows the breeder to choose genetically divergent parents for a breeding program. For example, it can be seen that increased heterosis could be obtained by crossing tall forms and dwarf forms. Interestingly, such crosses have been made in Sri Lanka between the form typica and forms dwarf yellow and dwarf green and the resulting hybrids out-performed both parents. It also appears that there is a narrow genetic base of coconuts in Sri Lanka. Therefore, importation of exotic germplasm for coconut breeding programme in Sri Lanka appears essential.

An extended study, which involved 75 tall individuals and 55 dwarf individuals representing 94 different coconut ecotypes from a world-wide collection also provided new insight and also evidence in support of previous hypotheses concerning the origin and dissemination of coconut. The results showed two main distinctive tall coconut groups: one comprising coconuts from Southeast Asia and the Pacific (hereinafter SAP coconuts) and another comprising coconuts from South Asian and African coconuts (hereinafter SAF coconuts). All dwarf types except a few of them grouped separately to form a sub cluster within the main-cluster of SAP coconuts. Interestingly, none of the dwarf types fell in the SAF coconuts. Lebrun et al. (1998) and Rohde et al. (1992) have reported very similar results. In both studies, dwarfs were found to cluster with SAP coconuts. Everard (1996) and Fernando and Rohde (1997) who surveyed a subset of Sri Lankan coconut using RAPDs and ISTR (Rohde et al., 1995); respectively, found that the tall coconut variety "San Ramon" (originating from the Philippines) grouped with Sri Lankan dwarfs when Sri Lankan talls coconut formed a separate cluster. Furthermore, dwarf coconuts showed a reduction in diversity and in number of alleles present when compared to the tall coconuts (42 alleles against 84). This suggests that dwarf coconuts are a subset of tall coconut and are directly evolved from tall coconut probably as a result of an event of domestication. Moreover; these results also suggest that dwarf coconut evolved from tall coconut originated in Southeast Asia and the Pacific; and, therefore somewhere in the Far East or the Pacific. The dwarf coconut found in Asia and East Africa are assumed to be early introductions whereas those in West Africa and America arrived only in the last 500 years (Harries, 1977) Dwarfness seemed to have occurred as a result of mutation as alleles for dwarfness seems to be rare in the tall coconuts even in the heterozygous state. Otherwise, the natural occurrence of dwarf palms in tall coconut populations would be a common phenomenon. It is likely that the selfing of heterozygous tall resulted in dwarf homozygotes and permitted the phenotypic expression of dwarfness together with the other recessive alleles, for example bright fruit colours. This is possible because tall coconuts are not strictly allogamous and the degree of selfing is likely to vary in natural populations. It has been observed that some coconut forms from Southeast Asia and Pacific showed a considerable degree of autogamy (Anuradha Upadhyay, CPCRI, India, personal communication). Since some recessive genes may be lethal, positive selection for recessive genotypes is necessary, for example susceptibility to soil moisture stress and pests results in dwarf coconuts being unable to survive in the wild. Thus selection for dwarfness may be automatically selected for inbreeding behavior and eventually fixation of this trait may have resulted in dwarf coconuts becoming predominantly in-breeders. Therefore, it can be hypothesized that a dwarf coconut is a collective result of mutation, self-pollination and an event of domestication.

SAP coconuts and SAF coconuts are genetically very distinctive (at the DNA level) based on SSR polymorphism. Harries (1978) too recognized this difference by phenotypic expression of the fruit components despite the variation due to environment factors. South Asian and African (both east and west) coconuts, therefore, seem to have originated from a common source and this result goes with Harries (1978). Panama talls (found in America) were found very closely related to South East Asian and the Pacific coconuts based on SSRs and this result also goes with Harrries (1978).

Harries (1978) suggests that SAF coconuts are the predominately naturally selected coconuts (‘Niu kafa’ type of coconuts) while coconuts in Southeast Asia are predominately domesticated.
coconut (‘Niu vai’ type of coconuts), from SAP coconuts, domesticated in the pre-agriculture era by early man for increased nut water as a source of fresh uncontaminated drinking water and the introgressed coconuts between those two types. The domesticated form was subsequently carried via sea voyages by the Polynesians and they introgressed with existing wild type populations as suggested by Harries (1998). Thus, coconuts in the Pacific includes ‘Niu kafa’ type, ‘Niu vai’ type and introgressed type (Harries, 1978). In order to confirm this hypothesis, further studies are necessary. DNA relationships observed in this study, however, show that coconuts found in Southeast Asia and the Pacific are genetically more close; and, in contrast, they are genetically quite different from what is found in South Asia and Africa. This is very clearly evident in the resulting dendogram, where coconut genotypes from Southeast Asia and the Pacific form a common cluster in which genotypes from both geographic are intermingled when genotypes from South Asia and Africa form a separate cluster. Moreover, in this study (although the number of samples studied was inadequate to be conclusive), slightly higher genetic variation was observed for the Pacific while similar levels of variation were observed for Southeast Asia and South Asia. Therefore, based on result of this study, it might be argued that both coconuts types found in the two geographical regions are either separate domestication events or events of multiple origins that occurred in the Far East or Pacific or in the Indian Ocean region.

A future DNA-based study involving improved sampling from all coconut growing regions including isolated undisturbed and uninhabited islands in the Far East and Pacific where wild coconut are likely to be present and the analysis of fossil coconuts would help unveil the origin, domestication and routes of dissemination of coconuts. Use of DNA technology to identify the site of domestication for einkorn wheat is an example of such a study (Heun et al., 1997). Collaboration between different scientific disciplines (taxonomy, biology, archaeology as well as population and molecular genetics and plant breeding) and all the techniques available (morphological traits and genetic markers including nuclear, chloroplast and mitochondrial markers and measurement of seed to pollen flow) are required. Cytoplasmic markers allow examination of variation passed through the maternal lineage in contrast to nuclear genetics. Therefore, understanding of the past history of a species from the present day distribution variation by the use of cytoplasmic markers is attractive. As such, emphasis should also be given for development of cytoplasmic markers in coconut. Attempts to develop cytoplasmic markers to distinguish between types of coconuts to characterize coconut populations showed no variation in coconut at the cytoplasmic level (Perera, 1999).

The high level of genetic diversity observed in the Sri Lankan tall coconut population based on SSRs and AFLPs in this study and the observable morphological and quantitative variation in the field, indicates prospects for further improvement of coconut through selection. Variation can be observed for flowering time, nut size, nut number, number of nuts per bunch, number of bunches produced per year, kernel weight and percentage of oil as well as the degree of resistance to water stress. The CRIC 60 coconut variety produced at the seed garden in the Coconut Research Institute of Sri Lanka (CRISL) is a good example of improvement of coconut through selection for yield which surpasses the yield of ordinary tall coconut. However, emphasis should be given to long-term evaluation of data in selection of palms as significant variations of characters have been observed between harvests and between years. Moreover, controlled pollination between selected palms is required; otherwise, contribution of unselected palms, as pollen donors, will diminish the effect of selection. Marker assisted selection will, therefore, play an important role in identifying genotypes that carry desirable combinations of characters independent of environment and will accelerate the improvement of coconut through selection and breeding. Therefore, construction of genetic linkage map is desirable to provide the basis for correlating the genotype (genetic markers) to phenotype (qualitative and quantitative). Because coconut is a perennial crop with a long generation time, development of a suitable mapping population is a more difficult task. As coconut and oil palm both are diploid with identical chromosome numbers (2n = 2x = 32), it was proposed that coconut SSRs may be used to augment the oil palm linkage map and compare synteny between these two species (Mayes et al., 1997). In order to examine whether coconut SSRs are conserved between oil palm and coconut, SSRs developed in this study were used to amplify oil palm DNA. However, the SSRs did
not amplify across these species. SSRs have been identified in Expressed Sequence Tags (EST) in Norway spruce (P. Hedley, personal communication) and these EST derived SSRs may exhibit greater conservation between related species. These types of markers may be valuable in linkage map construction. However, ESTs have not yet been developed for coconut or oil palm.

Since tall coconuts are highly heterogeneous, a mapping population for coconut can be constructed by crossing genetically divergent parents (Pseudo test cross) that show considerable variation for morphology and quantitative traits or selfing F1 hybrid resulting from such parents. The resulting progeny can be used to construct a genetic linkage map. The Pseudo test cross approach is being used to develop a linkage map for Norway spruce (P. Hedley, personal communication). A major constraint to such an approach in coconut is the limited number of seeds produced by the mother palm that results in insufficient seed from one tree. This is further aggravated by the low rate of success in artificial pollination in coconut. Continuous pollination can be carried out over a period of years to produce enough seeds. But the age difference between individuals of the progeny would then complicate field evaluation of the progeny. From the result of the SSR study it was shown that dwarfs are highly homozygous and therefore, it may be possible to produce identical genotypes by selfing. In Sri Lanka, fourth generation selfing progeny of dwarf green can be found growing in an isolated seed garden. Potentially, these progenies represent identical genotypes and could be simultaneously crossed with pollen from a tall parent to generate large quantities of seed in a very short time. This approach has been used to develop a segregating coconut population in Tanzania by crossing twelve Malayan red dwarf palms with a heterozygous Rennell tall (originating from Pacific), resulting in 60 segregating individuals. Collaboration under the supervision of BUROTROP (Bureau for the Development of Research on Tropical Perennial Oil Crops) has been set up to construct a skeleton map using AFLPs and SSRs. To date, the level of polymorphism between the parents using AFLPs are very low with only ten out of sixteen AFLP primer combinations tested by the author, being polymorphic with average of only 2.6 polymorphic bands per primer combination. AFLP primers that generated polymorphism between Sri Lankan tall (originating from South Asia), Rennell tall parents. This stresses the need to screen parental material to produce polymorphic segregating progeny for map construction as it is essential to choose highly polymorphic parents before embarking on a mapping project. Using the SSRs developed in this study, nine out eighteen SSRs were observed to be polymorphic between parents. The results of the analysis of world-wide collection of coconut, as described previously, suggests crosses between dwarf and tall coconut from South Asia and Africa or a cross between tall coconuts from South Asia/Africa and Southeast Asia/Pacific would generate a more attractive segregating family as they are shown to be genetically more different and are morphologically distinct.

Collaboration between laboratories using SSRs is feasible as primers can be exchanged between laboratories. Moreover, recent developments in fluorescent labelling of SSRs and semi-automated detection allow high throughput screening of individuals. SSRs also have the added benefit in linkage map construction providing common reference (based on allele size) between different mapping studies to be used in combination with other markers which are comparatively difficult to reference between populations. Once a skeleton linkage map is constructed using SSRs, other types of markers such as AFLPs, SSCP, RAPD and ESTs can be added to saturate the maps.

The development of a linkage map for coconut would be a very useful tool to locate QTLs (e.g. flowering time, number of nuts etc.). Furthermore, the identification of mapped SSRs will add power to the creation of genotypic databases. Such databases when compared with phenotypic data will provide coconut breeders with an important resource to identify parents with complementary traits and genetic estimation of divergence between palms. In the short term, the use of genetic markers to characterise coconut germplasm and improve the efficiency of identifying parents for crossing is likely to represent the larger impact of molecular markers on coconut improvement. The deployment of molecular markers, in breeding remains a challenge but is obviously highly desirable for coconut. The cost involved in establishing the facilities for marker assisted selection is high and
also requires considerable automation coupled with sophisticated Laboratory Information Management Systems (LIMS). To achieve this goal will require sustained collaboration between research teams and continuous support from various funding agencies.

Acknowledgement

I acknowledge Prof. Wayne Powell, Dr. J R Russell and Dr. Jim Provan from the Scottish Crop Research Institute (SCRI), UK and Dr. W M U Fernando, Mr. J M D T Everard and Mrs. C Banadaranayake from the Coconut Research Institute of Sri Lanka for their guidance and support. I greatly acknowledge the comments and suggestions made by Dr. Hugh Harries, CICY, Mexico.

References:


Heun, M; Schafer-Pregl, R; Klawan, D; Castagna, R; Accerbi, M; Borghi, B & Salamini F 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278: 1312-1314


Powell, W; Morgante, M; Andre, C; Hanafey, M; Vogel, M J; Tingey, S V & Rafalski J A 1996a. The comparison of RFLP, RAPD, RFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding **2**: 225-238


Rohde, W; Kullaya, A; Rodriguez, J & Ritter E 1995. Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of *copia*-like 16RI repetitive elements. Journal of Genetics and Breeding **49**: 179-186


Russell, J R; Weber, J C; Booth, A; Powell, W; Sotelo Montes, C & Dawson, I K 1999. Genetic variation of *Calycophyllum spruceanum* in the Peruvian Amazon Basin, revealed by amplified fragment length polymorphism (AFLP) analysis. Molecular Ecology **8**: 199-204