# Statistical Genetics of Coconut: An Overview

*V. Arunachalam and K. Muralidharan*

## Introduction

Area expansion and replanting of unproductive gardens are part of the strategies followed for enhancing production and productivity in plantation crops. A large number of quality planting material is to be made available every year for this purpose. The short-term objectives of crop improvement programmes in plantation crops are therefore formulated to meet this requirement leaving limited scope for a detailed examination of the genetic properties of all the respective breeding populations. Lack of readily available tools for the analysis of experimental data from perennial crops also complements to this situation. The statistical procedures, data generation techniques, estimates of genetic parameters and other related aspects are discussed below.

## Genetic Diversity

Assessment of genetic divergence is important for any breeding programme. The realization of heterosis in coconut necessitates the selection of populations for diverse origin. Conventionally, a field trial consisting of various cultivars of interest may be laid out and proceed with evaluation. Such a trial with coconut will occupy large area and requires many years for its completion. The data from such a trial will be voluminous and all the characters need not be of equal importance. Also not all characters may be available for different trials at different locations. Some characters may be more conservative than others. The fruit components of coconut are less affected by the growth of palms and management and hence preferred by many workers (e.g., Rao and Pillai, 1982). Fruit components may also be of use in periodical assessment of field genebanks. Multivariate techniques are sensitive to the inter correlation among variables, the possibility of variable reduction need to be examined before employing such techniques (Muralidharan *et al.*, 1993). To avoid duplicate accessions in the gene bank, the divergence analysis is employed at the time of collection (e.g., Zizumbo-Villareal and Pinero, 1998; Kumaran *et al.*,, 1998).

A number of procedures are available to measure the genetic diversity which include the coeeficient of parentage, Shannon-Weaver diversity index, Jacquard's coefficient, Nei-Li's genetic distance, modified Roger's distance, Euclidean distance, Mahalanobis' generalized distance principal component analysis (PCA), metroglyph analysis and discriminant analysis. Following the construction of a similarity or dissimilarity measure, the populations have to be grouped for which many methods are in practice such as unweighted pair group mean analysis (UPGA), average link (AL) method, partition method (PM) etc. Some of these procedures were employed to assess the genetic diversity in different regions and is summarised in Table 1. No general recommendation on the suitability of a procedure can be made from the results of these studies.

Santos *et al.* (1997) suggested the use of Mahalanobis' distance for studying inter population variability and PCA to represent a set of individuals without preliminary grouping and Fisher's discriminant functions for studying within population variability. Besides the ambiguity in selection of a procedure for studying genetic divergence, problems are encountered while interpreting the results. Selection of parents for hybridization based on divergence analysis need not give disired results always. For example, the devergent types viz., Andaman Giant Tall and Laccadive Micro Tall, as obtained in a study at CPCRI, when hybridized did not give heterosis for seedling characters in the $F_1$. Another aspect of divergence analysis is the set of characters to be included in the study. In two of the in situ studies with fruit components as the set of characters clearly identified the distint types viz., Red Dwarf (Ovasuru, 1993) and Coco Gra Tall - a makapuno type (Kumaran *et al.*,,1998).

Questions are also raised on the reliability of divergence analysis in coconut (Raveendran, personal communication) as each palm of a cultivar will be a half sib and therefore the within population variation will be confounded with the genetic variation. This aspect need to be considered while fixing the minimum size of population to be required for attempting divergence studies.

Diversity in coconut is available in uninhabited or far off coasts and islands especially wild, self-sown natural stands. Due to inherent transport and communication constraints, most of these sites are either unvisited or little time is available during visit. Under such difficult situations, data recording could be restricted to a few important characters identified by Kumaran *et al.* (1998) such as length and breadth of leaflet, spikelet length, inflorescence stalk length, husk thickness or husk to fruit proportion.

Table 1. Procedures employed for the analysis of genetic diversity in coconut

| Location | Procedure | No. of populations | Reference |
|---|---|---|---|
| *In situ* diversity | | | |
| Papua NewGuinea | Eucledian dist./UPGMA | 78 | Ovasuru (1993) |
| Madagascar, Mauritius and Seychelles | PCA/PM | 15 | Kumaran *et al.* 1998) |
| Orissa (India) | PCA/PM | 10 | CPCRI (1998) |
| West Bengal (India) | PCA/PM | 11 | CPCRI (1998) |
| Mexico | Eucledian dis./UPGMA | 41 | Zizumbo-Villarrealand Pinero (1998) |
| Pacific Ocean Is. | Mahhattan metric/UPGMA | 29 | Ashburner *et al.* (1997) |
| *Ex situ* diversity | | | |
| IRHO, Ivory Coast | Canonical Analysis | 17 | N'Cho *et al.* (1993) |
| CPCRI, India | Mahalanobis' dist. /AL | 70 | CPCRI, communictaed |
| KAU,India | Mahalanobis'dist. /PM | 24 | Balakrishnan and |
| TNAU, India | Metroglyph | 23 | Ravindra *et al.* (1987) |
| Japan (Philippines data) | PCA/.PM | 39 | Sugimura *et al.* (1997) |

**Field Gene Bank and Evaluation Trials**

Establishment and maintenance of coconut gene bank in the field is a tedious task. The requirement of large area, time and monitoring for pests and disease make it difficult to have large number of palms per accession. For evaluation trials, 40-60 plants per accession per replication were suggested by Rao and Batugal (1996) and 30 plants by Santos *et al.* (1997). A database management software to access palmwise data of an experiment is being prepared by COGENT.

**Sampling procedure and size**

A suitable strategy for sampling populations has to be devised. Because, a breeder is confronted with problems of huge populations, fast rate of gene erosion detecting useful low frequency alleles and unpredictable future needs.

Namkoong (1988) recommends dispersed sampling in perennial crops to deal the problem of defining independent genotypes in age, reproductive and spatial heterogenity. Marshall and Brown (1975) in sampling a species of subdivided populations, one approach is to disperse the samples into N locations, with n genotypes per sampling site. If a single allele of interest exist at frequency q in only certain populations, and those populations exist at frequency p, then Q of a single site sample would be,

$Q = [ (1-p) + p(1-q)^{2n} ]$ and multi-site (N) sample with n genotypes

$Q = [ (1-p) + p(1-q)^{2n} ]^N$

where Q is the probability of saving the allele.

Santos *et al.* (1997) recommends coarse and fine grid sampling method for germplasm collection. One fruit sample per tree is collected at random for F.C.A . IPGRI /COGENT recommends a minimum of 30 (Talls) or 10 (Dwarfs) individuals per site. Collection of seed nuts from more sites rather than from more individuals in a site has also been advised. A sample of 200 nuts or 600 embryos per population is essential to represent a population. Santos *et al.* (1997) has given a simple formula to arrive at sample size (N) required at the desired confidence interval (C.I ) for a given CV (coefficient of variability %) as, N= 1.96*CV/CI.

**Estimates of Genetic Parameters**

Partitioning of variance to its causative components and expressing the relative contribution of each component are helpful to know about the genetics of quantitative characters. The variance components are often obtained through the conventional least squares approach for balanced data. In reality, seldom the data are balanced and variances homogeneous. Added to these are the problems arising from

the long generation interval and time requirement for completing the evaluation of successive crops. Estimation and comparison of genetic parameters at different stages of a breeding programme are therefore a difficult task in this crop. Nevertheless few attempts were made in different countries in this direction. Two approaches are commonly followed for genetics studies of quantitative characters viz., controlled matings and correlation between relatives.

**Mating designs**

Only limited attempts were reported on application of mating designs in crop improvement of coconut. In a diallel experiment with selected palms of cultivar in Sri Lanka, Fernando (1996) obtained the heritabilility of nut weight as 0.45. The results of this study suggested the use of intermediate performing parents for crossing programmes as they combine both additive and on-additive components of genetic variation. In India, using palms of relative homogenity, a diallel was tried with nine cultivars (8 talls; 1 dwarf) which helped in identifying the best general and specific combiners. However, assumptions for a diallel program (Hayman, 1954) such as homozygosity of parents, absence of linkage between characters under study and lack of meternal effects cannot be expected to hold good with coconut. This is because (I) economic traits like copra and tendernut water being the endosperm components will have two-third contribution from the maternal parent and (ii) some palms (e.g., pre-potent) may have close linkage for desired characters.

Most of the coconut tall types are cross pollinated. Therefore, palm wise combining ability studies have to be undertaken. Such studies are reported to be helpful in refinement of released hybrids in Ivory Coast. Another approach could be the application of North Carolina Designs. Ravindran (personal communication) suggests line x tester matings similar to Kempthorne's sire x dam mating design which will provide information on combining ability and other genetic components of variance from the analysis half- and full- sibs. At IRHO, tall x tall and dwarf x tall hybrid improvement programs are based on reciprocal recurrent selection of half-sib families (Burdeix, 1998; Burdeix *et al.*, 1993). More studies are required to work out the relative merits and demerits of different mating designs in coconut.

**Data on relatives**

Estimation of genetic components of variance from data on relatives is advocated in many crops. Parental and juvenile selection are two common practices in coconut cultivation and breeding. Selection of mother palm is mainly based on annual yield (number of nuts per palm). Other characters of importance for mother palm selection include the number of functional leaves, length of stem (Patel, 1937), spherical or semi-spherical crown, short and stout bunch stalks and high copra yield (Menon and Pandalai, 1960). Though the assessment of genetic variability of these characters has been handicapped by (i) inadequacy of comprehensive germplasm (ii) limited population size in the genetic stocks and (iii) lack of well laid out experiments for their comparisons (Bavappa and Nampoothiri, 1974), estimates of heritability of nut yield for few populations were reported (Table 2). No standard error is provided to these estimates. The standard error is a function of sample size. For instance, with regard to a population of 55 open pollinated progenies of 5 West Coast Tall palms, the regression estimate of heritability for average annual yield over a ten year period was obtained as $0.3 \pm 1.4$. Because of this large standard error, the estimate was found to be non-significant. The non-adequate

Table 2. Estimates of heritability of coconut yield

| Description of population | Method of estimation | Character | $h^2$ |
|---|---|---|---|
| A total of 57 open pollinated progenies (at Pilicode) of 14 West Coast Tall palms (at Kasaragod) (Lakshmanachar, 1959) | Offspring-Parent Regression | Average annual yield over 10 year period | 0.49 |
| i) A total of 93 self pollinated progenies of 12 WCT palms | -do- | Annual yield (a single year) | 0.24 |
| ii) 16 open pollinated progenies of 6 WCT palms (Nambiar *et al.*, 1970) | -do- | -do- | 0.71 |
| 540 progenies derived from 108 palms grouped into 6 yield levels ('lines') which were pollinated in 3 different ways ('testers') (Nambiar and Nambiar, 1970) | (1) Line X Tester ANOVA<br>(2) Parent-offspring Regression (high yield groups) | No. of nuts<br>-do- | 0.47<br>0.22 |

sample size is a major problem with many coconut field experiments.

It is known that the heritability of a character will be increased when determined on the average of several measurements (Jain, 1992). Corresponding to Lakshmanachar's (1959) estimate, the heritability of annual coconut yield can then be worked out as 0.27 and for the biennial average it is 0.37. The repeatability of these two characters was taken as 0.5 and 0.7 respectively as reported by Muralidharan and Vijayakumar (1999). In summary, the annual nut production in coconut cannot be considered as a highly heritable character. The genetic gain from the practice of 'mother palm selection' based on annual or biennial yield alone is therefore expected to be marginal. Further, unlike the 'mass selection' as practiced in annuals where the selected individuals are put together en masse for mating, in coconut, open pollinated progenies are produced. A consequence to this is that the value of heritability to be considered for working out the genetic gain has to be halved as selection is applied only for one of the parents. Palms yielding more than 80 nuts per year are usually selected. Another suggestion was to select only the best 10% palms in a population. There can not be much difference by following these two criteria as it was reported by many workers that only 8 to 10% palms yield more than 80 nuts in Kerala (Satyabalan, 1984). Now with a moderate phenotypic selection differential as 40 nuts and heritability as 0.3, the expected genetic gain would be 6 nuts per generation. Dividing it by the generation interval as 15 years, the genetic gain per year will be 0.4 nuts per palm per year.

As mentioned earlier, selection is also practiced at the seedling stage. Nampoothiri *et al* (1975) studied the relationship of seedling characters with yield and found that selection could be based on the number of leaves and girth at collar. The genotypic correlation of these two characters with yield were reported to be 0.53 and 0.41 respectively. Mathew and Gopimony (1991) worked out the heritability of different seedling characters such as number of leaves (0.65), girth at collar (0.76), leaf area (0.76), height of seedling (0.63) and age at leaf splitting (0.63).

When multiple measurements of a character are available, the partitioning of variance corresponding to 'repeatability' too has some practical implications as the estimates of other genetic parameters are often not available. When a rectangular lay out of large number of trees (planted at random) are available, Pearce (1956) put forward a different method of partitioning the total variance into components due to the 'positional' and the rest, by including the coefficient of soil heterogeneity in the model. This approach was followed by Shrikhande (1958) and Pankajakshan (1960) to partition the variability of average annual coconut yield over even number of years into genetic and environmental. Muralidharan and Vijayakumar (1999) further partitioned the environmental variation into 'special-' and 'general-' environmental and attempted a non-linear fit to obtain the combined estimates of the parameters. The repeatability estimator thus obtained will be denoted as $r_{P\text{-}S}$, the suffix is to read Pearce-Shrikhande.

The increase in nut yield on advancement of the bearing age (Rao, *et al.* 1978; Muralidharan *et al.*, 1993) and the biennial pattern of bearing are the two major systematic variations associated with the coconut yield. The systematic variation due to the former cause may be removed by converting the yield of second and later years of bearing to the basis of first year (Muralidharan *et al.*, 1993). The problems arising from bienniality is generally overcome by taking average yield over even number of years or by converting the annual yields to first order moving averages as Wahi (1994) suggested. He also proposed an estimator of repeatability.

Smith *et al.* (1998) has developed a program PLANTVAR for rapid estimation of broad based heritability based on data of genotypes planted in bulk without progeny testing. This program could also be of very good help in coconut. Using the data available on many accessions at one location or few accessions at many locations heritability of many important agronomic, economic, physiological, and biochemical characters can be worked out. Earlier studies were restricted to single accession or few accessions for a few economic characters or colour markers.

**Genotype-by-Environment Interaction**

The analysis of data from multilocational trials in coconut is not straightforward. Only few cultivars may be common to all the trials. Also there may exist cultivar x year interaction in these trials, as agroclimatic conditions are known to have influence on nut yield. A common error in the analysis of cultivar x year interaction is the choice of an incorrect factorial model with year as one of the factors. Such an incorrect model was used by Patil *et al.* (1991) while analysing cultivar x year interaction in coconut (Muralidharan *et al.*, 1998). A correct approach could be based on the covariance structure of annual yields. However, under randomly assigned conditions (time), a split-block model is adequate provided, the compound symmetry properties (i.e. equal variances and equal covariances) hold good. The split-block anova is different from that of a split-plot for one additional source of variation viz. replication x time. The assumption of compound symmetry may be tested by using Box's criterion. On violation of this assumption, alternative methods are to be followed; the easiest being the

multivariate analysis. The general multivariate model may not always be estimable or efficient especially when the covariance of successive observations depends on the times of measurement and the number of measurement times large. Consequently more parsimonious models like 'random effects' or autoregressive (AR) models for the covariance structure are need to be considered. Application of rank stability statistics for comparison of sensitivity of cultivars over years may be seen in Muralidharan *et al.* (1993).

**Problem of WideVariation**

There are three levels of variation one has to encounter in coconut breeding viz., intra-individual (one palm may give variable number of nut/bunch in an year due to seasonal differences etc.), Intra-population (within a population, palms may differ in bearing capacity))and Inter-population. Populations belonging to different cultivars differ in many characters, which could be genetic or environmental, or interaction of both. Suitable methods to study the above mentioned three levels of variation could help to get clear interpretation of the results.

Bienniality in bearing and fluctuation in yield of nut or copra from year to year are major problems in comparing the performance of hybrids / varieties of coconut. To overcome the bienniality, CPCRI uses the mean yield of two successive years and for tackling year to year fluctuations cumulative nut yield over a period of 10 -15 years.

To measure the bienniality of a palm, use of Hublin's index seems to be an ideal method. It is the average of ratios of absolute difference between adjacent years to the corresponding sum

$1/n\ [^{n}{}_{j=1}\ Y_j - Y_{j+1} / (Y_j + Y_{j+1})]$

where Y is the yield recored during n number of consequetive years and $_j$ being the serial number of year when years are arranged in ascending order). Differences in intra/inter cultivar pre-bearing age and advancement of bearing years with increase in nut production pose problems in direct comparison of cultivars in initial years. To overcome this problem, Muralidharan *et al.* (1993) suggested correction of the yield by converting the yield of second and later years of bearing on first year basis.

Rao and Pillai (1982) have suggested the use of ratios instead of absolute figures for overcoming the problem of high CV. Normalisation of yield data attempted by CPCRI reveals the use of $(x+10)^{1/2}$ transformation in reducing the skewness and kurtosis. Calibration of a palm using an integrated index with principal component scores of vegetative characters as developed at Sri Lanka (Mathes *et al.*, 1996) could also be of help for dealing the problem of high CV.

Sometimes we confront practical problems like death of palms in a particular plot. Use of missing plots or nearest neighborhood analysis could help us here. In cases where a plant in the neighborhood dies or a plant grows better than other plants of same genotypes in the same plot because of a favourable environment like its proximity to water source, the difference could be analysed by nearest neighborhood analysis.

Ledig (1974) gives a formula for analysing the ratio of the responses to selection ($R_{ct}$ is the response at control and $R_t$ being the response of treatment) as

$R_{ct}/R_t = (1-r)\ [(n-I)/\ n(1-t)]^{1/2}$.

Here n is the number of trees used for comparison, r the degree of relationship between the candidate tree and those around it and t is the intra-class correlation coefficient measuring similarity within groups.

The method developed at CPCRI for predicting the annual yield (Y) of coconut based on total count of nuts above fist size $X_1$ and that of below fist size $X_2$ present in the crown at the time of observation could be useful in many cases.

| | | | |
|---|---|---|---|
| Rainfed : | Sep. to Feb. | $Y = 1.09X_1 + 0.44X_2$ | ($R^2 = 0.92$) |
| | Mar. to May | $Y = 1.02X_1 + 0.52X_2$ | ($R^2 = 0.88$) |
| Irrigated : | Sept. to Feb. | $Y = 1.09X_1 + 0.42X_2$ | ($R^2 = 0.92$) |
| | Mar. to May | $Y = 1.08X_1 + 0.42X_2$ | ($R^2 = 0.82$) |

**Future Thrust**

The major problem with respect to statistical analysis in coconut germplasm evaluation and field trials is that these analyses have been done either exclusively by statisticians or by plant breeders with poor knowledge of biometrics. Further, the experimental methods and interpretations have been analogous to that of annual crops which are not exactly applicable to perennial crops having long pre-bearing phase and purity of the planting material is doubtful due to the allogamous nature. So coconut requires statistical techniques which are tailor made for the crop. Therefore a lot of work has to be done this regard. A mannual developed based on the contributions of experienced coconut breeders and biometricians of different countries could be useful tool for coconut researchers.

**References**

Ashburner, G.R., W.K. Thomson., G.M. Halloran., M.A. Foale., 1998. Gen. resources & Crop Evol. 44:327-335.

Balakrishnan.P.C. and K.M.N. Nambodiri. 1987. Indian Coconut J. 40 : 9-17.

Bavappa, K.V.A. and K.U.K. Nampoothiri. 1974 Indian J.

Genet. 34(A): 58-65.

Bourdeix, R., A. Sangare., J.P. Le Saint., J. Meunier., J. P. Gascon., F. Rognon. and M. de Nuce de Lamothe. 1993. pp 79-94 In: Advances in Coconut Research and Development. Oxford & IBH, New Delhi.

Bourdeix, R. 1988. Oleagineux. 43 (7):283-295.

CPCRI .1998. Annual Report.

Fernando, WMU .1996.Cocos. 11:1-6.

Hayman, B.I. 1954. Genetics. 39 : 789-809

Jain, J. P. 1992. Statistical Methods in Quantitative Genetics. Hindustan Pub. Corpn., Delhi.

Kumaran, P.M., P.K. Koshy., V.Arunachalm., V.Niral. and V. A. parthasarathy. 1998. PLACROSYM XIII (In Press).

Lakshmanachar, M.S. 1959. Indian Cocon. J. 12: 65-68.

Ledig, F.T. 1974. Forest Sci. 20:2-16.

Mathew, T. and R. Gopimony. 1991. In: Coconut breeding and management, KAU, Vallanikkkara. pp 99-102.

Liyanage, D.V. and K.I. Sakai. 1961. Journal of Genetics. 57: 245-252.

Mathes, D.T., H.A.J. Gunathilaka., and D.N.S. Fernando. 1996 .Cocos. 11: 53-68.

Menon, K.P.V. and K.M. Pandalai. 1960. The Coconut Palm - A monograph. ICCC, Ernakulam.

Muralidharan, K. and K. Vijayakumar. 1999. J. Plant. Crops 27: 149-151

Muralidharan, K., M. J. Ratnambal., P. Gopalasundaram., and J. Mathew. 1998. J. Plant. Crops 26: 20-24.

Muralidharan, K., M. J. Ratnambal., and S. Dhamodaran. 1993. J. Plant. Crops (Suppl.) 21: 395-398.

N'Cho. Y.P, A. Sangare, R. Bourdeix, F. Bonnot and L. Baudouin. 1993. Oleagineux. 4893):121-132.

Nambiar, M.C. and K.P.P. Nambiar. 1970. Euphytica 19: 544-551.

Nambiar, M.C., J. Mathew, and Sumangalakutty. 1970. Indian J. Genet. 30: 599-603.

Namkoong, G. 1988. Hort. Science. 23 (1): 79-81.

Nampoothiri, K.U.K., K. Satyabalan., and J. Mathew. 1975. Phenotypic and genotypic correlations of certain characters with yield in coconut. FAO Tech. Wkg. Pty. Cocon. Prod. Prot. & Processing, Kingston, Jamaica.

Ovasuru, T. (1993). ACIAR Tech Bull., 53 : 33-40

Pankajakshan, A. S. 1960. Indian Coconut J. 13 37-43.

Patel, J.S. 1937. Proc. Assoc. Econ. Biol. 5: 1-16.

Patil, J.L., P.M. Haldankar., B.M. Jamadagni. and M.J. Salvi. 1993. J. Maharashtra Agricl Universities. 18(3) 361-364.

Patil, J.L., P.M. Haldankar., B.M. Jamadagni. and P. Rethinam. J. Plantn. Crops. 19:37-40.

Pearce, S. C. 1956. J. Indian Soc. Agri. Statist. 7 23-24

Rao, E.V.V.B. and R.V. Pillai. 1982. PLACROSYM. V: 112-124

Rao, E.V.V.B.,, K. Satyabalan., and J. Mathew. 1978. In: All India High Level Symposium Researches in Plant Cytogenetics, Waltair, 9-12 Feb., 1978. pp 15-16.

Rao. R.V and Batugal, P. 1996. (eds.) Proc. Regional Coconut Gene Bank Planning Workshop. IPGRI/ COGENT.

Raveendran, T.S. , T. Ramanathan., G. Nallathambi. and H. Vijayaraghavan. 1987. Cocos.5: 32-38.

Santos *et al.* (1997) (ed.). Stantech manual for coconut breeding. IPGRI/COGENT .

Smith, S.E., R.O. Kuel., I.M. Ray., R. Hui. and D. Soleri. 1998. Crop Sci. 38 : 1125-1129.

Satyabalan, K. 1984. Indian Cocon. J. 15: 3-11

Shrikhande, V. J. 1958. Indian Coconut J. 11 pp 140-157

Sugimura, Y., M. Itano., C.D. Salud., K.Ostuji., H. Yamaguchi. 1997. Euphytica. 98(1/2): 29-35.

Wahi, S. D. 1994. J. Ind. Soc. Agric. Statist. 46: 279-289.

Zinumbo-Villareal, D. and D. Pinero. 1998. Am J. Bot. 85:855-865.