*Cord* 2006, **22** (2)

# An effective population size for reliable map resolution of coconut (*Cocos nucifera* L.)

C.K. Bandaranayake[1]

**Abstract**

Size of the mapping population is a critical landmark in producing a genome map of any crop. In coconut there is difficulty in producing a reasonably large mapping population within a specified period of time. Therefore revealing an effective population size for standard map resolution of coconut is needed. It is found that the size of a segregating population with 400 individuals is efficient for consistent map resolution of coconut.

**Key words:** Coconut, genome mapping, JoinMap, map resolution, population size.

[1]Genetics and Plant Breeding Division, Coconut Research Institute, Lunuwila, Sri Lanka.
bandaranayakeck2000@yahoo.com

## Introduction

The coconut palm (2n = 32), which is popularly known as 'Tree of Life', has been used by human beings for at least half a million years as a source of food, drink and oil. It is socially, culturally and religiously associated with millions of people around the world especially Asia, Africa and Pacific Islands. Apart from being a source of healthy food and drink, it provides shelter, health, wealth and aesthetic sense. It not only provides sustainable income to millions of people who are directly and indirectly depending on this crop but also helps to sustain household food and nutrition security and alleviate poverty. Additionally, the oil derived from coconut contains about 49% lauric acid, a potential source for producing an antiviral, monolaurin, which is regarded as highly efficient in treating HIV/AIDs infected patients. Therefore, importance has been assigned to coconut genome mapping in order to detect the locations of genes on particular chromosomes, understand the level of inheritance of agriculturally important traits and identify molecular markers associated with those traits. The information derived from the study will prove to be of immense use in enhancing the socioeconomic status of the small and marginal coconut farmers spread over the producing countries.

Genome mapping needs a multi generation pedigree with a substantial segregating progeny. A mapping population should be fairly large to contain all genetic information from many segregating gametes (Helentjaris, 1988). The appropriate population size is very important when considering the map resolution and the order of the markers along a chromosome. Resolution of the map and the order of markers are critical milestones in genome mapping and they largely depend upon population size (Waugh, 2002). The main problem associated with coconut genome mapping is the difficulty in producing a reasonably large mapping population within a given period of time. It is mainly due to some inherited characters of coconut like limited number of flowers and nuts per tree per year, long time taken to develop into mature nuts, out-breeding nature and the large physical structure of the tree. Therefore, it is imperative to standardize an optimum size of the population for coconut genome mapping.

Simulation studies can be used to avoid the scarcity of resources. They can identify promising pathways and solutions that can sometimes direct research in to new lanes that are most likely to succeed. These simulation studies can act as an alternative for experiments, which are impossible, too dangerous or too costly to perform in the actual field. They can also be helpful to refine the experiments or even to make predictable suggestions. Two methods, which are quite different from each other, are used in genetic systems for simulation studies. One method is classical mathematical approach to population genetics and the other is creating individual model organisms and introducing them to behave in a way analogous to the behavior of real organisms. The basic idea involved in mathematical models is that the genetic properties of a population can be described in terms of relevant variables (gene frequencies) and parameters (e.g. selection pressures) and that the relationship of these may be defined by algebraic expressions. In such a way, the course of any change in the genetic constitution of the population may be calculated.

The computers, accurate and fast calculating machines well adapted to simulate the genotypes and genetic behavior of organisms, can be used for the operation of genotype simulation technique. But computer simulation is a slow process as compared with determinist**c** algebraic simulation. The other limitation for computer simulation is the fact that the individuals of the population have to be stored which will demand large storage space consistent with the size of the population. Computer based simulation techniques are of immense importance to study situations that are difficult to explore in realistic conditions. Simulation studies allow cautious interpretation of relatively complex genetic comparisons that have not been previously possible (Edwards and Page 1994) and sometimes they are important to investigate new paths (Crosby, 1973).

Genetic map for any crop would be very useful in order to gain better knowledge of the genome organization, to depict the genetic basis of key quantitative traits and to develop marker assisted selection for those traits. But the major constraint in genetic mapping of perennial tree crops is the unavailability of appropriate mapping populations. This limitation is more pronounced in coconut with its uncontrollable physical structure and breeding behaviour. Therefore the main objective of this study is to simulate several different sizes of progenies through special computer software and thus to select the effective population size for coconut genome mapping.

## Materials and methods

Seven different sizes of mapping populations, consisting of 50, 100, 200, 400, 600, 800 and 1000 individuals as progeny members, were simulated using the RiceSim software written in Fortran language by Prof. M.J. Kearsey, School of Biosciences, University of Birmingham, UK. Ten markers were assigned to each chromosome and altogether 160 markers for 16 chromosomes of coconut were placed.

Parental lines were produced via RiceSim software using highly homozygous dwarf coconut as mother parent ($P_1$) and highly heterozygous tall coconut as pollen parent ($P_2$). The important assumptions made at each simulation for each progeny were ten markers per chromosome, 0.1 recombination frequency (RF) between two markers placed on the adjacent loci in the same chromosome, 0.5 RF between two adjacent markers on different chromosomes and two alleles of equal frequency present in each marker. RiceSim was able to produce desired mapping populations successfully according to the given instructions. The results were written to an output file generated automatically by RiceSim and was placed on the same folder. Each file was automatically named accordingly as pop400.out was the output file for the population which contains 400 progeny members. RiceSim successfully generated the output files for all selected population types and were named as pop50.out, pop100.out, pop200.out, pop400.out, pop600.out, pop800.out and pop1000.out.

The locus genotype files for each population (Fig. 1) were prepared from the above output files as the input data files to be recognized by JoinMap software separately. Separate modules perform each task of JoinMap package. The first module after creating the raw data file is JMGRP that takes the responsibility in assigning markers to different number of linkage groups based on LOD scores already provided. The user has to decide which particular grouping is suitable to tally with the number of chromosomes. The module JMSPL split the defined grouping in to separate files corresponding to different linkage groups. The next step is to calculate pair wise recombination frequencies and modified LOD scores derived from the probability values of the chi-square test for independent segregation by JMREC module. The results of these calculations are written in to a “pair wise data file” per linkage group. Finally these pair wise data files are used as input files to the map construction module JMMAP, which produces a particular number of linkage groups. JMMAP requires a number of parameters such as the mapping function whether it is Haldane's or Kosambi's, the LOD threshold, the jump threshold, the triplet threshold, the ripple value, the number of top linkages, etc. to be specified from a response file. The map is a subsequent marker is added to a set of markers

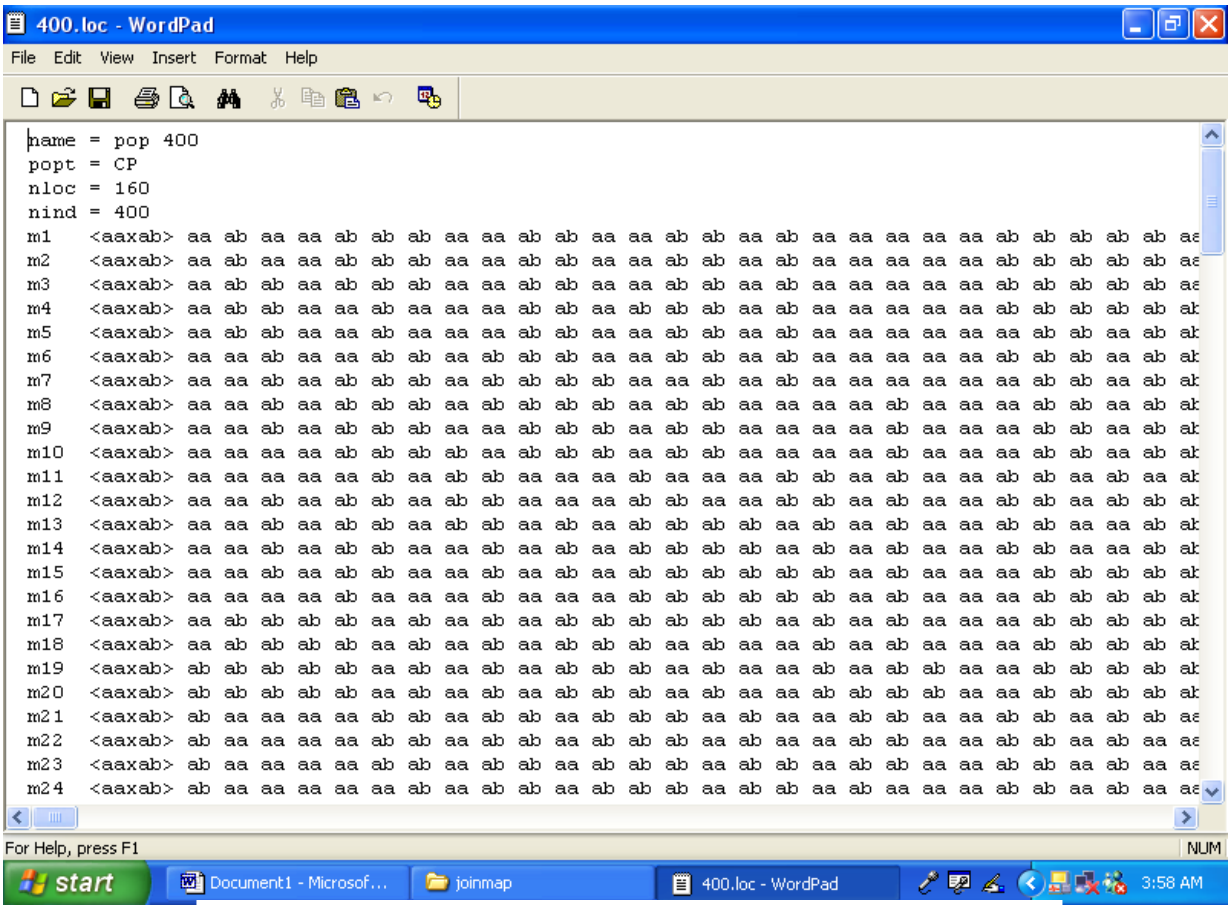


**Fig. 1. Locus genotype file for pop400 (pop400.out)**

for which the best fitting map has been calculated. Addition of a marker is done by a combination of trial-and-error placing it in a number of positions in the map region where the marker is likely to fit. JMMAP produces the output as a jmo-file that contains separate "top down" and "bottom up" lists of markers and map positions. These lists may serve as input files for the map-drawing program, DrawMap.

## Results and discussion

RiceSim was able to produce desired mapping populations successfully according to the given instructions. A significant difference was noticed between linkage groups of the population 50 and the rest (Table 1). The characters mentioned in Table 1 were common to all chromosomes except for first, fifteenth and sixteenth chromosomes. The marker order of the chromosomes shows gradual stabilization with the increasing size of the population as 100, 200 and 400 (Table 2 and Table 3). This is obviously noticeable in all linkage groups excluding one, six and sixteen. This irregularity could be explained by the parameters that used through long run of JoinMap software. These parameters include mostly the various LOD values and the phase of recombination between markers along each chromosome. Since the mapping populations were cross-pollinated ones, some of the parameters are beyond the processor's control or in other words the processor is clueless to alter them. Therefore the above anomalies could be observable within acceptable limits.

**Table 1. Differences between linkage groups of the population 50 and the rest**

| Character | Population size of 50 | Population sizes of 100-1000 |
|---|---|---|
| No. of linkage groups produced | 15 | 16 |
| No. of markers per linkage group | Less than or more than 10 except chromosomes 1, 15 and 16 | 10 |
| Marker order in linkage groups | Always not in order except 15$^{th}$ chromosome | Not in order in populations with 100 and 200 individuals but the rest was in order |
| Length of chromosomes | 4 chromosomes were around 50cM, 3 were above 120cM and the rest was around 100cM | All the chromosomes were around 100cM |

**Table 2. Marker order and distances in cM along linkage group 2 in populations of different sizes**

| Pop 50 | | Pop 100 | | Pop 200 | | Pop 400 | | Pop 600 | | Pop 800 | | Pop 1000 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m11 | 11.6 | m12 | 7.0 | m12 | 2.8 | m11 | 0.0 | m11 | 0.0 | m11 | 0.0 | m11 | 0.0 |
| m15 | 13.3 | m11 | 26.0 | m11 | 16.7 | m12 | 7.8 | m12 | 6.8 | m12 | 9.3 | m12 | 11.0 |
| m16 | 23.8 | m13 | 29.6 | m13 | 26.9 | m13 | 22.4 | m13 | 19.7 | m13 | 20.8 | m13 | 25.0 |
| m17 | 54.7 | m14 | 53.8 | m14 | 41.2 | m14 | 31.8 | m14 | 27.9 | m14 | 33.6 | m14 | 36.8 |
| m12 | 56.9 | m15 | 63.6 | m15 | 52.5 | m15 | 46.6 | m15 | 43.8 | m15 | 47.3 | m15 | 48.0 |
| m13 | 71.8 | m16 | 74.5 | m16 | 67.8 | m16 | 53.5 | m16 | 50.6 | m16 | 52.8 | m16 | 56.4 |
| m14 | 78.7 | m17 | 83.6 | m17 | 79.0 | m17 | 64.5 | m17 | 59.3 | m17 | 61.3 | m17 | 65.7 |
| m18 | 83.0 | m18 | 88.5 | m18 | 86.3 | m18 | 76.3 | m18 | 71.6 | m18 | 74.8 | m18 | 79.0 |
| m19 | 89.6 | m19 | 99.0 | m19 | 95.9 | m19 | 89.1 | m19 | 80.8 | m19 | 85.8 | m19 | 89.3 |
| | | m20 | 120.0 | m20 | 110.2 | m20 | 101.8 | m20 | 93.6 | m20 | 102.7 | m20 | 109.0 |

**Table 3. Marker order and distances in cM along linkage group 3 in population of different sizes**

| Pop 50 | | Pop 100 | | Pop 200 | | Pop 400 | | Pop 600 | | Pop 800 | | Pop 1000 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m20 | 0.0 | m21 | 0.0 | m21 | 0.0 | m21 | 0.0 | m21 | 0.0 | m21 | 0.0 | m21 | 0.0 |
| m21 | 4.7 | m22 | 5.3 | m22 | 9.2 | m22 | 10.8 | m22 | 11.8 | m22 | 12.0 | m22 | 11.7 |
| m22 | 11.0 | m23 | 19.0 | m23 | 14.7 | m23 | 20.3 | m23 | 26.0 | m23 | 27.2 | m23 | 27.3 |
| m26 | 21.4 | m24 | 35.9 | m24 | 33.2 | m24 | 31.6 | m24 | 35.5 | m24 | 39.1 | m24 | 35.0 |
| m27 | 30.7 | m25 | 46.5 | m25 | 36.8 | m25 | 48.0 | m25 | 50.1 | m25 | 51.9 | m25 | 50.4 |
| m28 | 46.0 | m26 | 56.8 | m26 | 50.4 | m26 | 55.0 | m26 | 60.5 | m26 | 63.5 | m26 | 63.6 |
| | | m27 | 66.1 | m27 | 69.8 | m27 | 68.8 | m27 | 71.5 | m27 | 77.1 | m27 | 72.2 |
| | | m28 | 76.6 | m28 | 74.6 | m28 | 80.5 | m28 | 82.8 | m28 | 85.2 | m28 | 80.9 |
| | | m30 | 82.9 | m29 | 82.0 | m29 | 92.3 | m29 | 93.7 | m29 | 96.1 | m29 | 93.4 |
| | | m29 | 87.6 | m30 | 90.0 | m30 | 100.8 | m30 | 103.9 | m30 | 105.3 | m30 | 104.0 |

When comparing the final chromosomal maps of different populations (from population size 50-1000), it is understandable that it moved towards to stabilized framework map with the increase of the size of the population. But it is important to observe that the marker order and

their locations along the chromosomes have not changed greatly ahead of the population size 400. This phenomenon is undoubtedly exposed on chromosome 2 and 3 in Tables 2 and 3 respectively. Though this is expected to be in whole genome, it is clearly visible in almost all chromosomes aside from chromosome one and six. That inconsistency may account for above mentioned genetic parameters that have been used through JoinMap software. On top of these observations regarding a cross pollinating crop with 16 pairs of chromosomes locating 10 markers each, the population size of 400 individuals is a crucial milestone to consider as an effective size for a mapping population.

The general belief is that the population size and the number of markers are equally important and critical milestones in genome mapping. It is often assumed that the more markers there are on a chromosome, the more precisely any quantitative trait loci (QTL) will be located. Simulations have shown that this is not generally true, four or five well-spaced markers are quite adequate, while further markers scarcely affect the precision (Hyne *et al*., 1995; Darvasi *et al*., 1993). It is also confirmed that the precision of QTL location depends more on sample size than on the density of markers and no great increase in precision is obtained with more than five well-spaced markers per chromosome (Kearsey and Pooni, 1996).

On top of the above facts, it is very unambiguous that the size of a mapping population would be a decisive part of information to create the power to give a firm image of the genome map. When the size of the mapping population increases, there is a possibility to access enough recombinants in different places of the genome. Therefore having a big population size, it is offered to have sufficient number of individuals comprising recombination at each and every point of the genome along all chromosomes. This is the principle of getting a stabilized map with increasing population size and that is the rationale for gradual stabilization of this particular map up to the size 400 individuals. However, it is always possible to alter the marker order and distances along a particular chromosome with inclusion of new markers, using different populations and changing the size of the same type of population. That is the explanation for slight changes beyond the population size 400 of this particular case. Though increasing size of population would help for reliable map resolution, there are practical problems to develop large mapping populations especially for crops like coconut. Therefore it is not meaningful to increase the size beyond this point because it only adds some inaccuracy due to the lengthily process of producing a big mapping population rather than increasing the precision of the resolution of the map.

## References

Bandaranayake, C.K. 2003. Investigation of the feasibility of constructing a map for coconut with several $F_2$ families using computer-simulated data. *Cord* **19(1):** 59-68.

Crosby, J.L. 1973. Computer simulation in Genetics. John Willey & Sons, London.

Darvasi, A., Weintreb, A., Minke, V., Weller, J. and Soller, M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**: 943-951.

Edwards, M.D. and Page, N.J. 1994. Evaluation of marker assisted selection though computer simulation. *Theor. Appl. Gene.* **88**: 372 – 376.

Helentjaris, T. 1987. A genetic map for Maize based on RFLP. *Trends Genet.* **3**: 217–221.

Hyne, V., Kearsey, M.J., Pike, D.J. and Snape, J.W. 1995. QTL analysis: unreliability and bias in estimation procedures. *Mol. Breed.* **1**: 273-282.

Kearsey, M.J. and Pooni, H.S. 1996. The Genetical Analysis of Quantitative traits. Stanley Thornes (Publishers) Ltd., UK.

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.* **3**: 739-744.

Waugh R., Dear, P., Powell, W. and Machray, C.G. 2002. Physical Education- new technologies for mapping plant genomes. *Trends Plant Sci.* **7**:521-523.