INVESTIGATION OF THE FEASIBILITY OF CONSTRUCTING A MAP FOR COCONUT WITH SEVERAL F_2 FAMILIES USING COMPUTER-SIMULATED DATA

By

***C. K. Bandaranayake*[1/]**

ABSTRACT

A computer simulation was performed using RiceSim computer software to explore the practicability of combining several different F_2 populations together through JoinMap to mimic the real available coconut mapping populations, and found that it was very successful. JoinMap would be able to map all 16 chromosomes which covered the map length of 1540 cM except for a single marker on chromosome 8. The largest marker interval was 32 cM at the bottom of chromosome 3 and all other markers were evenly distributed along the chromosomes maintaining the space around 12-30 cM between them.

INTRODUCTION

Coconuts consist mainly of two varieties, Tall and Dwarf, and they are genetically distinct for several important characters due to their different pollination behaviour. Tall type coconuts are generally cross pollinators while Dwarf type coconuts show mostly self-pollination. Hence, dwarf type coconuts are normally homozygous for many loci while tall type coconuts are highly heterozygous.

Breeding coconuts for high productivity and better adaptability to drought, pests and diseases is a major research priority in most of the coconut growing countries. For a long time, coconut breeding involved testing of inter-varietal and intra-varietal hybrids arising from morphologically distinct populations. However, the success of conventional breeding procedures is constrained by several palm characters such as the long juvenile period, out-crossing and heterozygous nature. In addition to that, the unavailability of a viable vegetative propagation mechanism and also the low genetic variability has caused many problems to the breeder. More recently utilization of molecular markers to enhance the coconut breeding work was initiated. Construction of a genetic map saturated with molecular markers would allow selection for characters to be carried out much more efficiently and effectively than conventional methods.

1/ Coconut Research Institute, Lunuwila, Sri Lanka.

The most critical decision in constructing a linkage map with DNA markers is the mapping population. A mapping population should be fairly large to contain all genetic information from many segregating gametes, but the currently available coconut populations are rather small to use for a satisfactory mapping programme. The reason for the small size of families is the production of limited number of seeds (nuts) from a particular mother palm within a fixed period of time. This is further aggravated by the low rate of success in artificial pollination in coconut. The development of inbred lines from heterozygous palms is almost impossible because of the long time taken for seed production in coconut.

Within all these constraints, the only available populations in coconuts are full-sib families with very small family sizes, half-sib families with few members obtained from controlled pollination using pooled pollen and several small F_2 families. If there is a possibility to merge several different small F_2 families together, a mapping population of practical size can be produced. Therefore it is advantageous, if we can combine small F_2 families for analysis coming from several parents that are genetically similar.

For most mapping projects, the most widely used genetic mapping software is Mapmaker (Lander *et al*., 1987). But the main disadvantage of that software is its inability to merge maps. JoinMap 2.0 (Stam, 1993) has been developed as an alternative to facilitate integration of genetic maps. When using JoinMap, at least some segregating markers must, obviously, be common to maps to be integrated. Integrated maps have already been produced for *Arabidopsis* (Hauge *et. al*., 1993), Barley (Qi *et. al*., 1996) and *Brassica oleracea* (Sebastian *et al*, 2000), where few maps were integrated and markers were common to at least two independent data sets. But considering the small F_2 families, it is worthwhile to merge several maps to maximize the population size. In order to explore the reliability of this approach, it was decided to use a simulation study.

Computer based simulation studies can be used to simulate situations that are difficult to explore in practice. It has allowed for tentative interpretation of relatively complex genetic comparisons that have not been previously possible (Edwards & Page, 1994). But there were few occasions in the past where the results obtained from such simulations investigate new pathways possible for some studies (Crosby, 1973; Sampson, 1984). Therefore simulations can be used as follows.

- To avoid commitment of scarce resources. In this case, even when it cannot replace any part of the actual laboratory work, simulation can often identify the most promising solutions and channel research into those paths most likely to succeed.
- An alternative for experiments those are impossible, too dangerous, or too costly to perform in the laboratory.
- Investigators control over the inputs to the system that may not be possible in the laboratory situation.

- To make predictions testable in the laboratory or suggest refinements for experiments. This case is very important for coconut and it is better to have results similar to these that might be gained from real situations because coconut has rather complex genome.

On top of the above facts, the objective of this study was to test the feasibility of constructing a map for coconut joining nine F_2 populations, exactly similar to the real situation, using computer-simulated data. The results obtained would inform the design of the laboratory experiment to be carried out for mapping coconut genome.

MATERIALS AND METHODS

Mapping population and markers

The actual coconut mapping population, for which seeds had been produced in Sri Lanka, consists of nine different F_2 families coming from self-pollination of nine different hybrid (Tall x Dwarf F_1) palms. To obtain this mapping population, it is needed to simulate P_1 (highly heterozygous tall coconut) and P_2 (highly homozygous dwarf coconut) and, from them, to simulate a large F_1 family. The parental lines were produced using RiceSim computer software, written in Fortran by Prof. M.J. Kearsey, School of Biosciences, University of Birmingham.

A population of size 1000 in linkage equilibrium was first simulated under the following assumptions.

- 6 markers on each of the 16 chromosomes (Table 1).
- Each marker had 2 alleles of equal frequency.
- Markers were 20 cM apart (i.e. each chromosome was 100 cM long).

Nine individuals were randomly selected from that large population to represent the nine possible F_1 hybrid plants. When selfing these randomly selected F_1 individuals to produce F_2 populations, free recombination was hypothesised between chromosomes (Recombination Frequency (RF) = 50%), but RF was 20% between each pair of adjacent loci. Each F_1 was selfed to produce F_2 populations having family sizes ranging from 40 to 80 (Table 2) to mimic the real families to be used for coconut genome mapping.

Setting up a map using JoinMap software

All F_2 populations were examined separately to produce data files for the JoinMap software programme. Different, separate locus genotype files were developed for each F_2 family using the notations displayed in Table 3. When 2 gametes were the same, those individuals were entered as homozygotes (AA for gamete type 1; BB for gamete type 2) and at times where 2 gametes are different those individuals were entered as heterozygotes (AB). Finally, each locus genotype file consisted of 3 different genotypes such as A (homozygous for female parent or AA), B (homozygous for male parent or BB) and H (heterozygous for both parents or AB).

Having produced correct locus genotype files, the next step in JoinMap programme is to assign markers to groups based on temporary pair-wise data files. Since JoinMap programme was unable to perform this,

the loci were grouped manually by neglecting monomorphic markers from each F_2 population separately. All other successive steps of JoinMap, such as splitting each linkage group to produce ordered linkage groups and finding recombination frequencies between each pair of markers to produce pair-wise data (PWD) files, were run separately for each F_2 until the PWD files formed. Nine PWD files were brought together within each group between nine F_2 populations and mapping was performed using joined PWD files for each linkage group.

RESULTS AND DISCUSSION

The JoinMap program was allowed to assign markers to linkage groups but it was unable to group them properly because the number of genes read from the locus genotype file was not equal to the number of genes read from temporary pair-wise data file due to some monomorphic markers. Therefore, grouping was performed manually, deleting monomorphic markers. The linkage groups of each F_2 family are illustrated in Table 2. When considering linkage group 1 of family F_2 (1), only 3 markers (M1, M2 and M6) were polymorphic out of six assigned as in Table 1. According to Table 2, all six markers of linkage group 1 were polymorphic in F_2 (2), F_2 (4), F_2 (8) and F_2 (9) but it was completely different in F_2 (5) and F_2 (6) having no polymorphic markers at all. The distribution of polymorphic markers between other linkage groups in each F_2 family is illustrated in Table 2.

Homozygous loci have occurred in F_1 individuals due to similar type of gametes received from grand parents at some loci (Table 3a). If we take gamete 1 that would have come from the grandmother and gamete 2 from grandfather, 3 different genotypes could be distinguished (Table 3b) in F_1 as AA (gamete 1=1 and gamete 2=1), BB (gamete 1=2 and gamete 2=2) and AB (gamete 1=1 and gamete 2=2). Considering two gametes in each F_1, it is easy to identify which markers will be monomorphic (AA or BB) and polymorphic (AB) in each F_2. According to Table 3, marker3 and marker4 would be monomorphic in F_2 coming from F_1 (1) because of the same gametes and thereby presence of homozygous loci in F_1 itself. Therefore those monomorphic markers would be unavailable for mapping at this stage because there is no variability. Because we used a population in linkage equilibrium for our F_1's with equal allele frequencies, almost half of the markers were monomorphic in each F_2 family.

JoinMap was unable to produce groups because it read only the polymorphic markers. Thus 16 groups were built manually deleting monomorphic markers so some groups were not accessible under some F_2 populations as illustrated in Table 2 (ex: groups 3, 11, 13 & 16 in F_2 (1) population). Missing groups have resulted from all markers being monomorphic in a particular chromosome. The condition would become worse when splitting and creating ordered groups. At this stage, there should be at least two markers on a group to retain it for further steps because no pair-wise interactions could be investigated with one marker per chromosome. Based on group files (Table 2), JoinMap did the splitting and then some groups disappeared when there was only one marker per group. Therefore, groups 4, 6, 9 and 12

will be lost in ordered groups of F_2 (1) (Table 4) in addition to those groups already missing, namely 3, 11, 13 and 16 in Table 2.

As a result of that, only about 8-10 linkage groups were available in each F_2 for further mapping as shown in Table 4. Actual PWD files were formed at the next step, JMREC, keeping LOD to a minimum and REC to a maximum. The JMREC was carried out and it defined suspect estimates for each linkage group based on given LOD threshold 0.001 and REC threshold 0.499. Table 5 shows the final linkage map according to the Kosambi mapping function by JMMAP. Based on suspect estimates resulting from JMREC, the final map would be restricted to a limited number of markers that are fully compatible and accurate.

Because the F_2 populations came from Tall x Dwarf hybrids, based on the results of Table 5, JoinMap would be able to map all 16 chromosomes except for one marker on chromosome 8. This covered a map length of 1540 cM. Only one locus (M48) was removed by JoinMap analysis due to conflicts within the linkage groups. The largest marker interval was about 32 cM at the bottom of chromosome 3. All other markers were evenly distributed along the chromosomes maintaining the space around 12-30 cM between them. The sequence of markers on each chromosome was correct.

Predictions and Refinements based on this study

- Integration of nine small F_2 populations can be done successfully, providing that there are at least six markers per chromosome with equal allele frequencies. This could and should be determined at the outset.
- Identification of markers that are polymorphic in each F_1 is vital before using them for F_2 populations, in order to reduce the wastage of time, energy and money.
- It is essential to use many markers to genotype the populations in order to make the process powerful and thereby to develop coverage of all linkage groups. In this case, it is very useful to use AFLP markers in addition to SSR markers to acquire many polymorphic markers for final mapping.
- The other option to make the process powerful is to increase the number of F_2 populations. Hence, many F_2 populations can be combined through PWD files so that all markers would be polymorphic at least in several F_2 populations. Therefore no marker would be removed at early stages of JoinMap program.

ACKNOWLEDGEMENT

Author wishes to thank Prof. M. J. Kearsey, School of Biosciences, University of Birmingham for providing RiceSim computer software and unlimited support in this project.

REFERENCES

CROSBY, J.L .1973. Computer Simulation in Genetics. John Wiley & Sons, London.

EDWARDS, M.D. and PAGE, N.J. 1994. Evaluation of marker assisted selection through computer simulation. *Theor. Appl. Genet.* **88**: 376-382

HAUGE, B.M., HANLEY, S.M., CARTINHOUR, S., CHERRY, J.M. and GOODMAN, H.M. 1993. An integrated genetic /RFLP map of the *Arabidopsis thaliana* genome. *The Plant J*. **3** :745-754

LANDER, E.S., GREEN, P., ABRAHAMSON, J., BARLOW, A., DALY, M.J., LINCOLN, S.E. and NEWBURG, L. 1987. MAPMAKER: an interactive computer programme for constructing genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174

QI, X., STAM, P. and LINDHOUT, P. 1996. Comparison and integration of four barley genetic maps. *Genome* **39**: 379-394

SAMPSON, J.R. 1984. Biological Information Processing: Current Theory and Computer Simulation. A. Wiley-Interscience Publications, John Wiley & Sons, New York.

SEBASTIAN, R.L., HOWELL, E.C., KING, G.J., MARSHALL, D.F. and KEARSEY, M.J. 2000. An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. *Theor. Appl. Genet*. **100**:75-81

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

Table 1. Assigned markers for each linkage group.

Chromosome	Markers					
1	M1	M2	M3	M4	M5	M6
2	M7	M8	M9	M10	M11	M12
3	M13	M14	M15	M16	M17	M18
4	M19	M20	M21	M22	M23	M24
5	M25	M26	M27	M28	M29	M30
6	M31	M32	M33	M34	M35	M36
7	M37	M38	M39	M40	M41	M42
8	M43	M44	M45	M46	M47	M48
9	M49	M50	M51	M52	M53	M54
10	M55	M56	M57	M58	M59	M60
11	M61	M62	M63	M64	M65	M66
12	M67	M68	M69	M70	M71	M72
13	M73	M74	M75	M76	M77	M78
14	M79	M80	M81	M82	M83	M84
15	M85	M86	M87	M88	M89	M90
16	M91	M92	M93	M94	M95	M96

Table 2: Polymorphic markers identified and associated with each linkage groups by Join Map in each of the nine F_2s. (Markers M1-M6 should be on Group 1, M7-M12 on Group 2 etc. The sizes of each F_2 family are indicated in brackets.)

Group	F_2 Families and their family sizes								
	$F_2(1)=80$	$F_2(2)=40$	$F_2(3)=50$	$F_2(4)=70$	$F_2(5)=40$	$F_2(6)=40$	$F_2(7)=50$	$F_2(8)=40$	$F_2(9)=70$
Group 1	M1 M2 M6	M1 M2 M3 M4 M5 M6	M1 M2 M4 M5	M1 M2 M3 M4 M5 M6			M2	M1 M2 M3 M4 M5 M6	M1 M2 M3 M4 M5 M6
Group 2	M9 M10 M11 M12	M7 M8 M9 M10	M7	M7 M8	M7 M8 M9	M7	M8 M9 M10	M7 M9 M10 M11	M7 M8 M10
Group 3		M16 M17 M18		M13 M14 M15 M16 M18	M14 M15 M16 M17 M18	M16 M17 M18	M13 M14 M15 M16 M17 M18		
Group 4	M19		M23	M19 M20 M21	M22		M19 M20 M21 M22 M23 M24	M23 M24	M24
Group 5	M28 M29 M30	M26	M25 M26 M27		M25 M26 M27 M28 M29 M30	M25 M26 M27 M28 M29 M30	M25 M26 M27 M28 M29 M30	M30	M26 M27 M28
Group 6	M36	M31 M32 M34 M35 M36	M31 M32 M33		M31 M32 M33		M32 M33 M34 M35		M35 M36
Group 7	M38 M41 M42		M39	M38 M39 M40 M41		M37 M38 M42	M39 M40 M41	M38	M37 M38 M39 M40 M41 M42
Group 8	M43 M44 M47	M46 M47	M43 M44 M45			M43	M43 M44 M45 M46 M47	M44 M45 M46 M47	M43

Table 2 continued…

Group	F_2 Families and their family sizes								
	$F_2(1)=80$	$F_2(2)=40$	$F_2(3)=50$	$F_2(4)=70$	$F_2(5)=40$	$F_2(6)=40$	$F_2(7)=50$	$F_2(8)=40$	$F_2(9)=70$
Group 9	M49			M52 M53 M54	M50 M53 M54		M49 M50	M49 M50	M49 M51 M52 M53 M54
Group 10	M55 M56	M56	M55 M56 M57 M58 M59 M60		M55 M56 M57 M58 M59	M55 M56 M57	M55	M55	M59
Group 11		M61 M62	M61			M61 M62 M63 M64 M65 M66		M61 M62 M63 M64 M65 M66	M61
Group 12	M68	M67 M68 M69 M70 M71 M72	M67 M68 M69 M70 M71 M72	M67 M68 M69 M70 M71 M72	M72	M67 M68 M69 M70 M71	M67 M68 M69 M70 M71 M72	M67 M68 M70 M71 M72	M69 M70
Group 13		M73	M73 M74 M75 M76 M77 M78	M73 M74 M75	M73 M74 M76 M77		M73 M78	M73 M74 M75 M76 M77 M78	M73
Group 14	M79 M80 M81 M82 M83	M79 M80 M81 M82	M79 M80 M81 M84	M79 M80 M81		M79 M80		M79 M81 M82 M83 M84	M80 M81 M82 M83 M84
Group 15	M85 M86 M87 M88 M89 M90	M85 M86 M87 M88 M90	M89 M90	M85 M86	M88 M89	M85 M86 M87 M88 M89 M90	M85 M86 M87 M88 M89	M87 M88 M89	M90
Group 16			M91 M92 M95 M96	M91 M92 M93 M94 M95 M96	M91 M92 M93 M94 M95	M92 M93 M94	M95 M96	M91 M95 M96	M91 M92 M93 M94 M95 M96

Table 3: Sample demonstration of the F_1 population.

Table 3a: Markers that have same gamete type are in bold under each individual.

F1	Gamete	Markers									
		1	2	3	4	5	6	7	8	……	96
1	1	1	2	**1**	**2**	1	1	2	2		2
	2	2	1	**1**	**2**	2	2	1	1		1
2	1	1	1	**2**	2	2	**2**	1	2		**1**
	2	2	2	**2**	1	1	**2**	2	1		**1**
3	1	**2**	1	**1**	1	1	1	1	**1**		2
	2	**2**	2	**1**	2	2	2	2	**1**		1
4	1	2	**1**	**1**	**1**	2	2	2	1		1
	2	1	**1**	**1**	**1**	1	1	1	2		2
:											
:											
9	1	1	1	1	2	2	1	1	2		**1**
	2	2	2	2	1	1	2	2	1		**1**

Table 3b: Genotype of each F1 under each marker based on the pattern of gametes in Table 4.3a.

F1	*Marker*									
	1	2	3	4	5	6	7	8	.....	96
1	H	H	A	B	H	H	H	H		H
2	H	H	B	H	H	B	H	H		A
3	B	H	A	H	H	H	H	A		H
4	H	A	A	A	H	H	H	H		H
:										
9	H	H	H	H	H	H	H	H		A

Table 4: Ordered linkage groups for each F_2 family.

F_2 Families from TxDG F_1 palms	Ordered linkage groups available
F_2(1)	1, 2, 5, 7, 8, 10, 14 and 15
F_2(2)	1, 2, 3, 6, 11, 12, 14 and 15
F_2(3)	1, 5, 6, 8, 10, 12, 13, 14, 15 and 16
F_2(4)	1, 2, 3, 4, 7, 9, 12, 13, 14, 15 and 16
F_2(5)	2, 3, 5, 6, 9, 10, 13, 15 and 16
F_2(6)	3, 5, 7, 10, 11, 12, 14, 15 and 16
F_2(7)	2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 15 and 16
F_2(8)	1, 2, 4, 8, 9, 11, 12, 13, 14, 15 and 16
F_2(9)	1, 2, 5, 6, 7, 9, 12, 14 and 16

Table 5: Final linkage groups with their marker positions from top to bottom.

Linkage Group 1		**Linkage Group 2**		Linkage Group 3		**Linkage Group 4**	
Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)
M1	0.0	M7	0.0	M13	0.0	M19	0.0
M2	20.0	M8	22.2	M14	19.7	M20	15.6
M3	35.0	M9	42.0	M15	35.9	M21	37.4
M4	55.7	M10	66.1	M16	55.5	M22	69.7
M5	76.5	M11	89.7	M17	71.6	M23	87.6
M6	98.5	M12	106.7	M18	103.3	M24	104.1

Linkage Group 5		**Linkage Group 6**		**Linkage Group 7**		**Linkage Group 8**	
Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)
M25	0.0	M31	0.0	M37	0.0	M43	0.0
M26	12.7	M32	21.8	M38	25.5	M44	22.5
M27	34.9	M33	44.7	M39	46.8	M45	41.1
M28	54.8	M34	67.2	M40	60.7	M46	67.8
M29	71.8	M35	96.3	M41	75.7	M47	85.6
M30	92.9	M36	117.1	M42	99.1		

Linkage Group 9		**Linkage Group 10**		**Linkage Group 11**		**Linkage Group 12**	
Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)
M49	0.0	M55	0.0	M61	0.0	M67	0.0
M50	16.3	M56	15.2	M62	18.6	M68	16.8
M51	38.5	M57	27.3	M63	41.9	M69	30.4
M52	58.8	M58	42.5	M64	59.2	M70	42.6
M53	83.4	M59	64.8	M65	72.9	M71	53.5
M54	97.0	M60	90.7	M66	97.1	M72	80.0

Linkage Group 13		**Linkage Group 14**		**Linkage Group 15**		**Linkage Group 16**	
Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)
M73	0.0	M79	0.0	M85	0.0	M91	0.0
M74	28.9	M80	21.4	M86	17.8	M92	20.8
M75	47.6	M81	36.1	M87	34.2	M93	36.2
M76	69.5	M82	54.6	M88	52.8	M94	52.1
M77	79.2	M83	76.6	M89	71.2	M95	65.9
M78	101.4	M84	90.6	M90	92.5	M96	82.5