Theor Appl Genet (2002) 104:1185–1191
DOI 10.1007/s00122-001-0820-1

P.A. Butcher · E.R. Williams · D. Whitaker · S. Ling
T.P. Speed · G.F. Moran

Improving linkage analysis in outcrossed forest trees – an example from *Acacia mangium*

Received: 22 January 2001 / Accepted: 05 June 2001 / Published online: 27 March 2002

Abstract Mapping in forest trees generally relies on outbred pedigrees in which genetic segregation is the result of meiotic recombination from both parents. The currently available mapping packages are not optimal for outcrossed pedigrees as they either cannot order phase-ambiguous data or only use pairwise information when ordering loci within linkage groups. A new package, OUTMAP, has been developed for mapping codominant loci in outcrossed trees. A comparison of maps produced using linkage data from two pedigrees of *Acacia mangium* Willd demonstrated that the marker orders produced using OUTMAP were consistently of higher likelihood than those produced by JOINMAP. In addition, the maps were produced more efficiently, without the need for recoding data or the detailed investigation of pairwise recombination fractions which was necessary to select the optimal marker order using JOINMAP. Distances between markers often varied from those calculated by JOINMAP, resulting in an increase in the estimated genome length. OUTMAP can be used with all segregation types to determine phase and to calculate the likelihood of alternative marker orders, with a choice of three optimisation methods.

Keywords Linkage analysis · Mapping · Outcrossed pedigrees · Codominant markers · Multi-locus likelihood · Optimisation

Communicated by P. Langridge

P.A. Butcher (✉) · E.R. Williams · G.F. Moran
CSIRO Forestry and Forest Products, P.O. Box E4008,
Kingston, ACT 2604, Australia
e-mail: Penny.Butcher@csiro.au
Fax: +61-2-62818233

D. Whitaker
Statistics Department, University of Waikato, Private Bag 3105,
Hamilton, New Zealand

S. Ling · T.P. Speed
Department of Statistics, University of California, Berkeley,
CA 94720-3860, USA

Introduction

There have been a number of approaches to the construction of genetic linkage maps. A maximum likelihood technique for map construction was developed by Lander and Green (1987) and implemented in the packages MAPMAKER (Lander et al. 1987), CRI-MAP (Green 1988) and MULTIMAP (Matise et al. 1994). The method simultaneously uses information from all markers in a linkage group to determine marker order. These programs were first developed for mapping human pedigrees where linkage phase was known and progeny sizes were limited. MAPMANAGER (Manly and Elliot 1991), which also uses multi-locus likelihood, focuses on mapping larger progeny arrays from inbred lines. JOINMAP (Stam et al. 1993) was developed to combine linkage data from different experiments or pedigrees. It uses only pairwise recombination fractions between markers to estimate marker order and therefore cannot be expected to be as accurate as multi-locus likelihood-based packages (reviewed by Ott 1992).

Genetic linkage maps for forest trees are generally constructed for outbred pedigrees where segregation is the result of meiotic recombination from both parents. Inbreeding depression has precluded the development of inbred lines, and long generation times have led to the use of two-generation pedigrees for mapping in which the linkage phase between pairs of loci is not known a priori. Unfortunately the likelihood-based packages (CRI-MAP, MULTIMAP, MAPMAKER and MAPMANAGER) are not optimal for mapping forestry species; in particular they cannot adequately accommodate phase-ambiguous data. JOINMAP can be used to map all segregation types but, as already mentioned, the underlying method is not optimal.

The added complexity of mapping outcrossed pedigrees using multi-locus likelihood has been addressed in two ways. In the double pseudo-testcross strategy, linkage is analysed separately for each parent (Grattapaglia and Sederoff 1994). This strategy has been used for the

analysis of segregation data from dominant loci in forest trees. However, the integration of maps remains problematic (Maliepaard et al. 1997). In addition, as most quantitative trait loci (QTL) in outcrossed pedigrees are likely to be multi-allelic, only codominant multi-allelic markers will permit the tracking, understanding and adequate manipulation of all the allelic variation segregating at QTLs (for example Byrne et al. 1997). Sewell et al. (1999) proposed a method for analysing codominant loci in outcrossed pedigrees where separate maps were first constructed for the male and female parents using multilocus likelihood. Phase was determined from the grandparents for the majority of loci. For phase-ambiguous loci, marker phenotypes were reciprocally coded for coupling and repulsion, and the correct phase determined based on the recombination fractions. Maps were then combined using JOINMAP. Neither strategy provides a one-step analysis for segregation data from outcrossed pedigrees.

In this paper we discuss a new program to overcome the limitations of existing packages and provide a phase-determining, likelihood-based procedure for linkage analysis in outcrossed forest trees. The features of a Windows package called OUTMAP, for carrying out the analysis, are described. Data from two outcrossed pedigrees of *Acacia mangium* Willd are used to compare results from OUTMAP with previously published results from JOINMAP.

Outline of OUTMAP

The theory and methodology for OUTMAP is described by Ling (1999). The implementation of this theory in OUTMAP is outlined below.

OUTMAP features

The package is Windows-based and has been written in C++. The user specifies a file of segregation data; this includes the parental genotypes followed by the progeny genotypes in a text file. An option can also be selected for data with known grandparental genotypes. The package automatically assigns a numeric code according to the parental segregation type; this is used when determining linkage phase. The package then generates linkage groups according to the specified maximum recombination fractions (maxrf) and minimum LOD score (minlods). The markers in the linkage group can be ordered using a selected optimisation procedure. The user can drop markers from a linkage group and also determine the likelihood of a fixed marker order. These options facilitate the comparison of OUTMAP results with those from other packages.

Segregation types

The package was developed for linkage analysis of segregation data from codominant markers. Four parental segregation types for loci are recognised.

1) Fully informative (FI) – heterozygous for both parents with the number of alleles equal to three or four.
2) Heterozygous (HE) – heterozygous for both parents with the number of alleles equal to two. With this type of locus, the package later analyses the offspring to determine whether the locus is heterozygous coupling or repulsion.
3) Female segregating (FS) – heterozygous for the female parent and homozygous for the male parent.
4) Male segregating (MS) – heterozygous for the male parent and homozygous for the female parent.

Linkage groups

The segregation type of each locus is assigned, and an analysis of each pair of markers is carried out to infer phase and estimate the recombination fraction and LOD score; all of these operations are carried out at a pairwise level. An algorithm is then used to form the linkage groups based on pairwise recombination fractions and LOD scores. The user defines values for maxrf and minlods. The definition of a linkage group is that for any locus in the group there is at least one other locus, such that the pairwise recombination fraction is less than maxrf and the corresponding LOD score is greater than minlods. Note that at this stage, pairwise information is sufficient to form linkage groups; in fact, most mapping packages agree closely on the formation of groups. For deciding the order of markers and estimating the distances between markers in linkage groups, pairwise information is no longer adequate.

Parental haplotypes

For each linkage group, OUTMAP uses an algorithm to determine the parental haplotypes for all of the markers in the group. This process uses the segregation type of each locus and if, as is common for data from outcrossed trees, grandparental genotype data are not available, offspring data are used. Details of the algorithm are given by Ling (1999). Note that this algorithm, to determine the parental haplotypes over all of the markers in a linkage group, is more sophisticated than the pairwise phase method used above to estimate recombination fractions for the formation of the linkage groups. At this stage the package will output the phase-determined data file, and the user has the flexibility to modify the file (e.g. override the automatic phase determination, drop markers) before the maximum likelihood stage.

Maximum likelihood

After linkage phase has been determined for all loci in a linkage group, the multi-locus likelihood can be evaluated for a particular ordering of the markers. Ling (1999) describes a variant of the Lander and Green (1987) algorithm that caters for the situation of outbred crosses with many offspring and possibly no grandparental data; the algorithm uses hidden Markov models (Rabiner 1989) and the EM algorithm (MacLachlan and Thiriyambakam 1996) to estimate recombination fractions. The process is complicated by the fact that the likelihood surface is non-convex.

Optimisation

Optimisation procedures are used to choose the marker order that maximises the likelihood. Three methods are available in OUTMAP including 2-opt and 3-opt, which were derived from the travelling salesman problem in which the salesman is asked to find the shortest route that connects a number of cities (Johnson 1990) and nested simulated annealing (Whitaker 1995). We have found that no single method is to be preferred for ordering markers. For example, 2-opt and 3-opt are often used in preference to the steepest descent part of the simulated annealing algorithm; the latter, however, usually deals better with the non-convex likelihood surface to finish off the search. The use of several optimisation methods can also provide a check on results.

Application of OUTMAP

Marker order and map distances were calculated for two unrelated, two generation, outcrossed, full-sib pedigrees of *Acacia mangium*, using the OUTMAP package. The pedigrees and markers are described by Butcher et al. (2000a, b). Segregation data were available for 252 co-dominant loci; 219 restriction fragment length polymorphism (RFLP) loci and 33 microsatellite loci. The number of markers of each segregation type are listed in Table 1. The log likelihood of marker orders and distances between markers in each linkage group from OUTMAP were compared with the previously published marker orders and distances based on JOINMAP (Butcher and Moran 2000). The marker orders in the JOINMAP-generated map were resolved only after considerable time was spent examining recombination fractions between pairs of markers to resolve inconsistencies between the two pedigrees and after comparisons were made with the order of mapped markers from MULTIMAP. In contrast, the OUTMAP order was taken directly from the package.

Table 1 Number of markers of each segregation type in two outcrossed pedigrees of *Acacia mangium*

Segregation type[a]	Cross A	Cross B
FI	34	51
HE	16	14
FS	60	61
MS	59	64
Total	169	190

[a] FI, Fully informative; HE, heterozygous for both parents with only two alleles; FS, heterozygous for the female parent only; MS, heterozygous for the male parent only

Results and discussion

Data handling

Parental and progeny genotypes of the two pedigrees had been scored concurrently with alleles numbered from one to eight according to size. These data were imported directly into OUTMAP which automatically assigned a numeric code according to the parental segregation type. This eliminated the step of recoding alleles alphabetically, according to five segregation types, which was necessary before using the JOINMAP package.

Linkage groups and marker order

The same linkage groups were formed using OUTMAP (maxrf = 0.3; minlods = 3) as were produced with JOINMAP, with three exceptions. Groups 4 and 5 in cross A and groups 13 and 14 in cross B (Butcher and Moran 2000) were combined in OUTMAP. These groups were, however, subdivided based on large recombination fractions between two markers; [cross A – recombination fraction = 0.480 between g749 (group 4) and g861 (group 5); cross B – recombination fraction = 0.385 between g90a (group 13) and g493 (group 14)]. The linking of groups 13 and 14 in cross B was, however, consistent with the marker order of an integrated map formed when data from the two pedigrees were combined (see Butcher and Moran 2000). The third exception was the marker g492, which was unlinked in OUTMAP but was mapped at the end of linkage group 9 in cross A using JOINMAP. This marker was a relatively large distance (29 cM Kosambi) from the adjacent marker and was noted as having an unusual segregation pattern (Butcher et al. 2000b). It was included in group 9, cross A, despite the large distance to the adjacent marker, as the marker order was consistent with that in cross B (Butcher and Moran 2000). One other marker was unmapped in cross A and two markers were linked to each other but not to other markers. This is consistent with reported results from JOINMAP (Butcher and Moran 2000).

The negative log likelihoods obtained from OUTMAP after optimisation in each linkage group are presented in Table 2. They are compared with the negative log likelihoods for the JOINMAP marker orders obtained by entering these fixed orders into OUTMAP. There were differences in marker order in half of the linkage

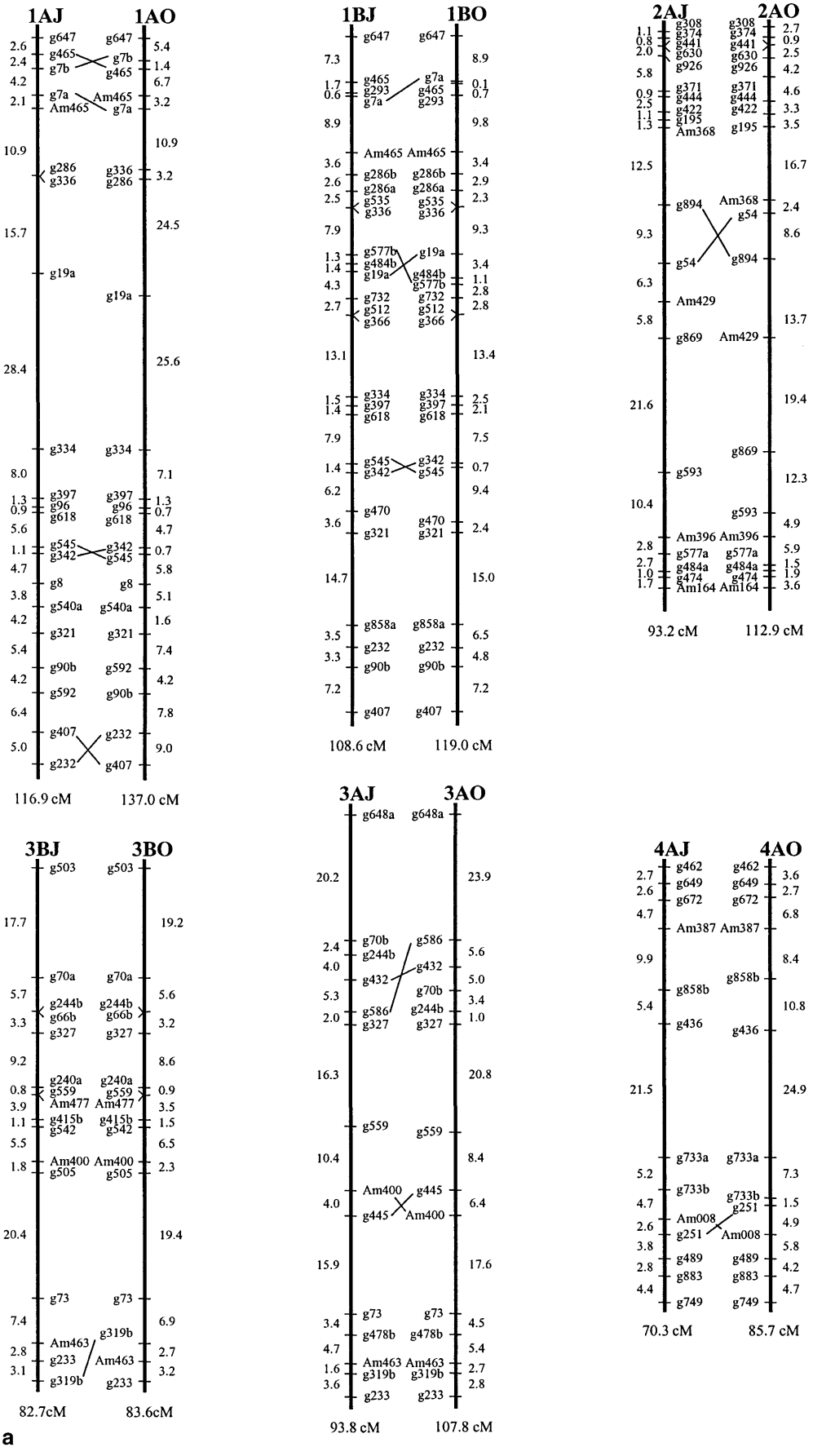
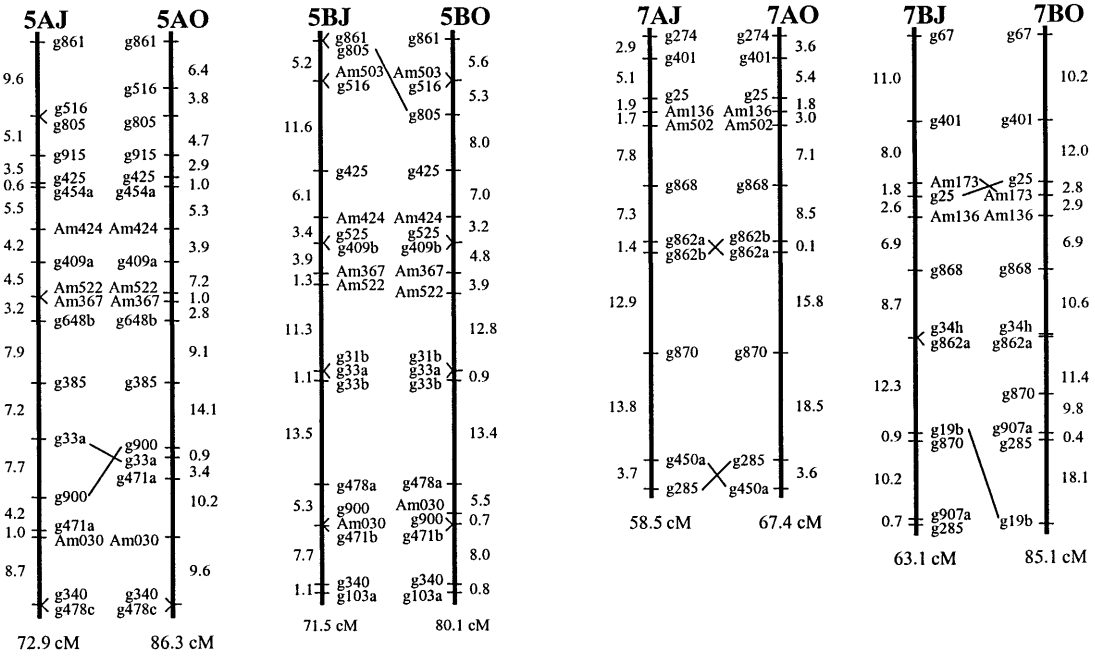
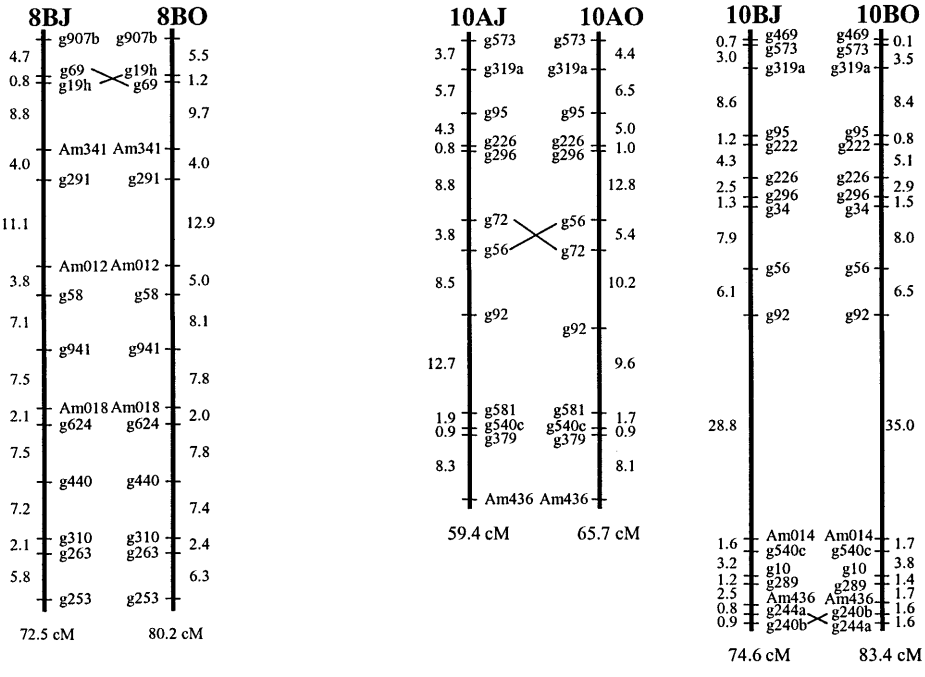
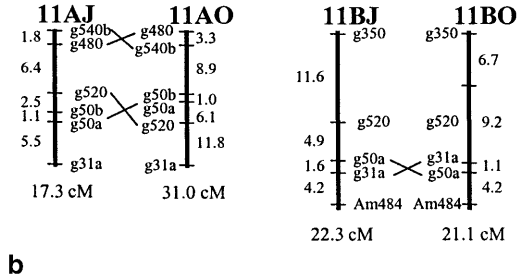


Fig. 1 Linkage groups mapped from two unrelated pedigrees (cross A and B) of *Acacia mangium* showing differences in map order and distances (centiMorgans Kosambi) produced using JOINMAP (*J*) and OUTMAP (*O*). Markers with the prefix *g* are genomic RFLP markers and those with the prefix *Am* are microsatellite markers. The OUTMAP groups are not drawn to scale


5AJ
5AO
5BJ
5BO
7AJ
7AO
7BJ
7BO
58.5 cM
67.4 cM
63.1 cM
85.1 cM
72.9 cM
86.3 cM
71.5 cM
80.1 cM


8BJ
8BO
10AJ
10AO
10BJ
10BO
72.5 cM
80.2 cM
59.4 cM
65.7 cM
74.6 cM
83.4 cM


11AJ
11AO
11BJ
11BO
17.3 cM
31.0 cM
22.3 cM
21.1 cM

b

Table 2 Comparison of negative log likelihood of marker orders generated by JOINMAP and OUTMAP for two independent linkage maps of *A. mangium* (= same order produced by both packages)

Linkage Group	Cross A		Cross B	
	OUTMAP (-log likelihood)	JOINMAP (-log likelihood)	OUTMAP (-log likelihood)	JOINMAP (-log likelihood)
1	872	889	1,029	1,054
2	875	887	899	=
3	753	779	680	698
4	666	670	711	=
5	721	721	782	794
6	554	=	699	=
7	516	517	655	666
8	646	=	733	735
9	193	=	479	=
10	490	494	713	726
11	302	305	319	322
12	320	=	696	=
13	215	=	313	=
14	191	=	293	=
15	115	=		

Table 3 Comparison of lengths (centiMorgans Kosambi) of linkage groups estimated from OUTMAP and JOINMAP

Linkage Group	Cross A	Length		Cross B	Length	
	Number of markers	OUTMAP (cM)	JOINMAP (cM)	Number of markers	OUTMAP (cM)	JOINMAP (cM)
1	21	137	117	26	119	109
2	20	113	93	17	113	101
3	14	107	94	16	84	83
4	13	86	70	12	80	77
5	18	86	73	19	80	72
6	11	76	69	15	90	84
7	11	67	59	12	85	63
8	15	83	71	14	80	73
9	3	10	8[a]	7	64	51
10	12	66	59	17	83	75
11	6	31	17	5	21	22
12	8	27	23	12	93	80
13	3	18	18	4	30	28
14	4	35	36	6	23	18
15	3	8	8			
total	162	950	815	182	1045	936

[a] Comparison excludes g492

groups in both pedigrees. The changes affected markers of all segregation types; 30% FI, 46% MS, 19% FS and 5% HE. The differences generally involved a change in the position of two loci, with a maximum of three changes in any one linkage group (Fig. 1). Where there were differences in order, the marker orders produced using OUTMAP were of higher likelihood than those produced by JOINMAP with the exception of linkage group 5, cross A, where the likelihood was the same for both orders (Table 2).

While the changes in the ordering of loci in the *A. mangium* maps were generally minor, they will affect the precision of QTL mapping as well as being important when comparing chromosome organisation among species. In the majority of cases (14 out of 24) the changes had no effect on the alignment of maps between the two pedigrees (cross A and B) because the markers concerned only segregated in one cross or the order was changed in both crosses. In five cases the changes in order improved the alignment of maps, while in the remaining five cases the changes resulted in differences in order between cross A and B.

Map distances

The distances between markers estimated using OUTMAP often varied from those estimated using JOINMAP. The differences were neither proportional nor in the same direction. However, the overall lengths of linkage groups estimated with OUTMAP were consistently higher than those from JOINMAP (Table 3). This resulted in total map length estimates from OUTMAP which were approximately 10% greater than from JOINMAP. A similar difference was reported between JOINMAP and MULTIMAP (Butcher and Moran 2000). The lower estimated map lengths from JOINMAP compared to multi-locus likelihood programs have been attributed to differences in the mapping algorithms and to assumptions concerning cross-over interference (Qi et al. 1996). In the *A. mangium* data, loci

were sometimes mapped to the same position in JOINMAP despite crossovers being evident in the progeny arrays. This can occur because JOINMAP adjusts map distances between a pair of markers based on the recombination fractions between other pairs of markers, using a weighted least squares analysis. It has also been argued that the multi-locus likelihood methods of Lander and Green assume an absence of interference, so where there is interference JOINMAP will correctly produce shorter maps, even when both programs use the Kosambi mapping function (Stam 1993).

OUTMAP successfully handles all segregation types, determines phase, provides a choice of three optimisation methods and can calculate the likelihood of alternative marker orders. It is specifically designed for analysing segregation data from codominant loci in outcrossed pedigrees and deals effectively with phase ambiguous data. The inherent risk of introducing errors when recoding data to suit the input format of different programs is avoided. In addition, there is no need to divide segregation data into separate data sets for male and female meiosis. OUTMAP is therefore recommended for mapping in all outcrossed pedigrees. Further details on use and distribution of the package are available from the second author.

References

Butcher PA, Moran GF (2000) Genetic linkage mapping in *Acacia mangium* 2. Development of an integrated map from two outbred pedigrees using RFLP and microsatellite loci. Theor Appl Genet 101:594–605

Butcher PA, Decroocq S, Gray Y, Moran GF (2000a) Development, inheritance and cross-species amplification of microsatellite markers from *Acacia mangium*. Theor Appl Genet 101:1282–1290

Butcher PA, Moran GF, Bell R (2000b) Genetic linkage mapping in *Acacia mangium* 1. Evaluation of restriction endonucleases, inheritance of RFLP loci and their conservation across species. Theor Appl Genet 100:576–583

Byrne M, Murrell JC, Owen JV, Kriedemann P, Williams ER, Moran GF (1997) Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. Theor Appl Genet 94:674–681

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121–1137

Green P (1988) CRI-MAP. Department Genetics, Washington University School of Medicine, St Louis, Mo.

Johnson DS (1990) Local optimisation and the travelling salesman problem. In: Paterson AH (ed) Automata, languages and programming. 17th Int Colloq Proc. Springer, Berlin Heidelberg New York, pp 446–461

Lander ES, Green P (1987) Construction of multi-locus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Ling S (1999) Constructing genetic maps for outbred experimental crosses. PhD thesis, University of California, Berkeley, Calif.

MacLachlan GJ, Thiriyambakam K (1996) The EM algorithm and extensions. John Wiley & Sons, New York

Maliepaard C, Jansen J, Oijen van JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. Genet Res 70:237–250

Manly KF, Elliot RW (1991) RI MANAGER, a microcomputer program for analysis of data from recombinant inbred strains. Mam Genome 1:123–126

Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MULTIMAP): a human genome linkage map. Nat Genet 6:384–90

Ott J (1992) Analysis of human genetic linkage, revised edn. John Hopkins University Press, Baltimore, pp 116–117

Qi X, Stam P, Lindhout P (1996) Comparison and integration of four barley genetic maps. Genome 39:379–394

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77:257–286

Sewell MM, Sherman BK, Neale DB (1999) A consensus map for loblolly pine (*Pinus taeda* L.) I. Construction and integration of individual linkage maps from two outbred three-generation pedigrees. Genetics 151:321–330

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant J 3:739–744

Whitaker D (1995) A nested simulated annealing algorithm. J Stat Comp Simul 53:233–241