Genet. Res., Camb. (2004), **84**, *pp.* 47–55. *With* 4 *figures.* © 2004 Cambridge University Press
DOI: 10.1017/S0016672304006974 Printed in the United Kingdom

The effect of an imprecise map on interval mapping QTLs

K. G. DODDS[1]*, R. BALL[2], N. DJOROVIC[2] AND S. D. CARSON[2]‡
[1] *AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand*
[2] *Forest Research, Sala Street, Private Bag 3020, Rotorua, New Zealand*

(*Received 12 August 2003 and in revised form 5 May 2004*)

Summary

The statistical analysis of quantitative trait locus (QTL) experiments relies on the use of a linkage map of the markers genotyped. Such a map is, at best, a good estimate of the true map. Resources might be diverted into developing better marker maps or improved maps become available after the analysis, raising concerns over the original analysis. It is therefore important to understand the sensitivity of QTL analysis to map inaccuracy. We have used simulation methods to investigate the consequences of an incorrect map on the results of a QTL analysis using interval mapping. Backcross data sets were generated with a particular map and then analysed with both the correct map and incorrect maps. If the incorrect maps maintained the true linkage groups (i.e. no markers were incorrectly assigned to another linkage group), the accuracy of the map had little or no impact on the ability to detect QTLs, the true significance levels of the tests or the relative placement of QTLs. When a marker was incorrectly placed on another linkage group, there was a small increase in the level of the test. After adjusting for this increase, there was a decrease in power to detect a QTL near the misplaced marker. This decrease was of a similar magnitude to that found when using a single-marker analysis compared with interval mapping. These results mean that QTL analyses can proceed without the need for very accurate marker maps, and that estimated QTL positions can be translated onto updated maps without the need for reanalysis.

1. Introduction

The advent of easily assayed sets of genetic markers spread across a genome has allowed systematic searches for quantitative trait loci (QTLs) – genomic positions associated with variation in quantitative traits (Botstein *et al.*, 1980; Bovenhuis *et al.*, 1997; Kearsey & Farquhar, 1998). 'Interval mapping' methods (Lander & Botstein, 1989), in which each position in the genome is considered for the presence of a QTL by examining the cosegregation of the nearest marker(s) with the trait, have become the favoured method of analysis. These methods allow the localization of a putative QTL with respect to a map. For crosses between inbred lines, they might provide a modest increase in the power (Darvasi *et al.*, 1993; Lander & Botstein, 1989; Rebai *et al.*, 1995). In outcrossing species, in which a marker might not be fully informative in the cross, information from nearby markers contribute to the analysis and so increase the power of the study (Bovenhuis *et al.*, 1997).

The application of interval mapping requires the use of a marker map, which has been estimated and therefore contains inaccuracies. Maps might already exist for populations other than the one under study (and markers might have been chosen on the basis of such maps). In this case, the question must be addressed of whether the previously published map (the 'standard' map) should be used or a new map should be estimated from the current data. Either of these maps could suffer to a greater or lesser extent from two types of error: sampling errors (influenced by sample size) and genotyping errors. Genotyping errors have been found to have a large effect on map accuracy, particularly for map length (Tomfohrde *et al.*, 1992). The standard map might have the advantage that genotyping errors have been minimized

* Corresponding author. e-mail: ken.dodds@agresearch.co.nz
‡ Present address: Carson Associates, 82 Tihi Rd, Rotorua, New Zealand.

through consistency checks with close markers, whereas the new map would usually have more widely spaced markers. However, the new map might be based on a different strain or pedigree or contain a different sex ratio of meioses to the standard map. In these cases, researchers might consider abandoning the standard map in favour of the new map. How important are these decisions?

Here, we investigate the importance of map accuracy in the analysis of QTL experiments. We wish to know which types of map error might have an effect. For example, there might be errors in map distances that are proportional (relative positions of markers correct), errors that are non-proportional but with correctly ordered markers or errors in marker placement (order or linkage group assignment). We also wish to know what sorts of effects such errors have on the critical level (to maintain the desired type I error rate), experimental power and bias in estimating map position.

2. Method

Data sets were simulated for five different 'true' models (QTL and map characteristics), shown as A–E in Table 1. These were chosen to cover a range of situations of interest when QTL mapping, including the absence of QTLs (A,D) or the presence of a single QTL (B,C,E), evenly spaced true maps (A–C) or maps with a higher density at one end (D,E), and centrally positioned QTLs (B) or QTLs towards one end but in the centre of a marker interval (C,E). In all cases, fully informative marker data were simulated on a single chromosome of length 100 cM (Haldane units) in a phase known backcross family using the RCross program from QTL Cartographer (Basten *et al*., 1994, 1997). A family size (n) of either 100, 300 or 1000 offspring was used. The simulated trait was normally distributed with a residual (excluding any QTL effect) standard deviation of 0·968. Where present, the QTL had an effect of 0·5 units, explaining 6·3% of the phenotypic variance, so that the phenotypic standard deviation was 1.

The simulated populations were then analysed using the 'true' map (map 1) and a series of incorrect maps (Table 2, Fig. 1). The incorrect maps were a compressed map (map 2), an expanded map (map 3), a map that is compressed at one end (map 4), a map that is expanded at one end (map 5), a map that is compressed at one end and expanded at the other (map 6), a map with one marker erroneously assumed to be in the same position as another (map 7), a map with two markers incorrectly ordered (map 8) and a map with one marker assumed to be unlinked, placed in the centre of another 100 cM chromosome (map 9). 1000 populations were simulated for each true-map–assumed-map combination. For incorrect maps in which a marker was assumed to be on another chromosome, an additional chromosome with only that marker was also included in the analysis (Fig. 1).

Table 1. *Characteristics of the 'true' map used in the simulation models*

Model	Number of QTLs[a]	QTL position	True map (Haldane cM)
A	0	–	10, 30, 50, 70, 90
B	1	50 cM	10, 30, 50, 70, 90
C	1	20 cM	10, 30, 50, 70, 90
D	0	–	5, 10, 25, 55, 85
E	1	7·5 cM	5, 10, 25, 55, 85

[a] Where a QTL was simulated, it was of size 0·5, accounting for 6·3% of the phenotypic variation in the backcross.

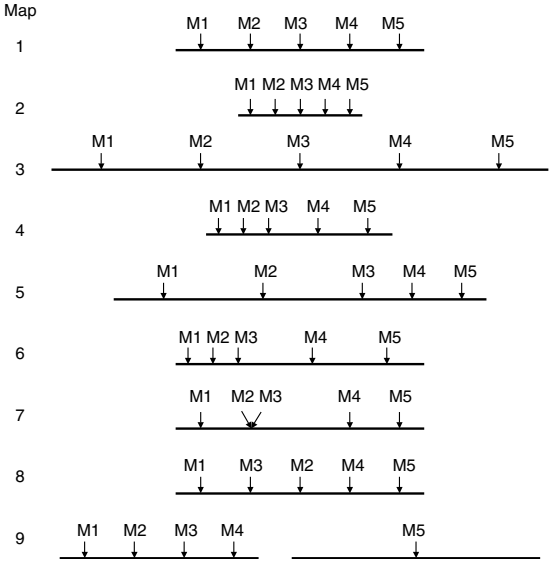

Fig. 1. Marker maps used for analysis of data generated using model A. Markers are labelled M1 to M5 according to their order on the true map. Map discrepancies for other models are similar in nature to the maps with the same number as shown here.

QTL Cartographer was used to analyse the simulated data, giving an interval-mapping likelihood-ratio test statistic for the QTL at its best estimated position (evaluated in 2 cM steps starting at 0·01 cM after each marker). The critical values corresponding to an experiment-wise type-I error rate of 5% was found as the 95th percentile of the test statistic in analyses in which no QTL was modelled and in which the true map was used (cases A1 and D1). These 5% thresholds varied between 6·23 and 6·71. For each case, the number of times the test statistic (at the best position) exceeded the relevant threshold out of the 1000 simulated populations was recorded. For each 'true' model (A–E), comparison of the powers to

https://doi.org/10.1017/S0016672304006974 Published online by Cambridge University Press

Table 2. *Assumed maps used to analyse data from each simulation and the power achieved using a 5% threshold*

Model[a]	Analysis map[b]	Assumed map[c] (Haldane cM)	Power[d]		
			$n=100$[e]	$n=300$[e]	$n=1000$[e]
A	1	10, 30, 50, 70, 90	0·050	0·050	0·050
	2	5, 15, 25, 35, 45	0·054	0·050	0·066
	3	20, 60, 100, 140, 180	0·056	0·040	0·062
	4	5, 15, 25, 45, 65	0·064	0·049	0·071*
	5	20, 60, 100, 120, 140	0·060	0·053	0·055
	6	5, 15, 25, 55, 85	0·051	0·053	0·066
	7	10, 30, 30, 70, 90	0·051	0·049	0·052
	8	10, 50, 30, 70, 90	0·055	0·065	0·066
	9	10, 30, 50, 70, U	0·057	0·038	0·064
B	1	10, 30, 50, 70, 90	0·086	0·133	0·372
	2	5, 15, 25, 35, 45	0·082	0·130	0·375
	3	20, 60, 100, 140, 180	0·074	0·135	0·373
	4	5, 15, 25, 45, 65	0·071	0·112	0·352
	5	20, 60, 100, 120, 140	0·085	0·127	0·353
	6	5, 15, 25, 55, 85	0·080	0·133	0·388
	7	10, 30, 30, 70, 90	0·075	0·145	0·355
	8	10, 50, 30, 70, 90	0·085	0·134	0·381
	9	10, 30, 50, 70, U	0·082	0·116	0·376
C	1	10, 30, 50, 70, 90	0·061	0·123	0·334
	2	5, 15, 25, 35, 45	0·079‡	0·094	0·324
	3	20, 60, 100, 140, 180	0·077	0·097	0·299‡
	8	10, 50, 30, 70, 90	0·080‡	0·107	0·330
	9	U, 30, 50, 70, 90	0·079‡	0·086	0·315
D	1	5, 10, 25, 55, 85	0·050	0·050	0·050
	2	5, 7·5, 15, 30, 45	0·053	0·060	0·061
	3	10, 20, 30, 70, 110	0·047	0·053	0·060
	8	10, 5, 25, 55, 85	0·057	0·049	0·052
E	1	5, 10, 25, 55, 85	0·065	0·148	0·336
	2	5, 7·5, 15, 30, 45	0·073	0·135	0·329
	3	10, 20, 30, 70, 110	0·081	0·129	0·344
	8	10, 5, 25, 55, 85	0·079	0·139	0·334

[a] Refer to Table 1 for the characteristics of the true map for each model.
[b] Map 1 uses the 'true' map. All other analyses use maps with errors.
[c] Marker positions in the order of the true map, with markers which are assumed to be unlinked denoted by U.
[d] Proportion of 1000 simulated populations with test statistics over the threshold.
[e] n refers to the number of simulated progeny.
* $P<0{\cdot}05$ compared with Map 1.
‡ $0{\cdot}05<P<0{\cdot}10$ compared with Map 1.

detect a QTL using an incorrect map (analyses 2–9) compared with that using the true map (map 1) represents the effect of using an incorrect map.

For each population, the estimated QTL position (relative to the assumed map) was recorded. We also calculated these estimated QTL positions translated to their positions on the true map. Translation was linear in map distance, mapping the position relative to the two flanking reference points (markers or end-points) on the assumed map to the same relative position on the true map. For example, if the QTL was one-third of the distance from the first to the second marker on the assumed map, it would be one-third of the distance between these markers on the true map. This holds regardless of whether the order of these markers is the same on the true and assumed maps, and of whether there are any intervening markers on the true map. We refer to these positions as 'translated positions' and distances between them as 'translated distances'. For cases A7 and B7, assumed position 30 cM translates to the region 30–50 cM on the true map. However, we only translate analysis positions, which do include this position. Fig. 2 illustrates the translations for models A and B. Where the 'true' model included a QTL, the distance between the estimated and true QTL positions was also calculated on this translated scale. Translated positions were not calculated where one marker was assigned to another chromosome, because the linear arrangement of markers was then lost.

https://doi.org/10.1017/S0016672304006974 Published online by Cambridge University Press

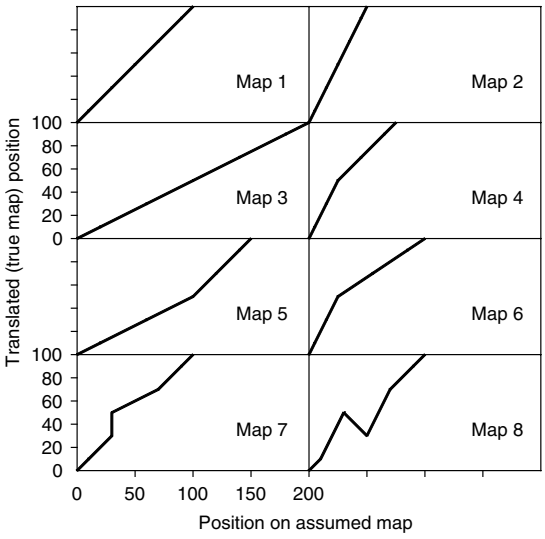

Fig. 2. Trellis plot of translated positions corresponding to positions on assumed maps 1–8 used with models A and B.

3. Results

The power of these simulated experiments are the proportions of the time that the observed test statistics exceeded the corresponding threshold (chromosome-wise 5% level using the correct map) and therefore the proportion of the time that the hypothesis of no QTL present is rejected at this level. These powers are shown in Table 2. For cases in which there was no QTL modelled (models A,D), these values were around 0·05–0·06 for all population sizes simulated and all analyses. These values estimate the chromosome-wise type I error. Analyses 1 results are exactly 0·05 by definition of the thresholds used, and comparison of this value with those from the other analyses with models A and D portray the effect of incorrect maps on the significance level.

For models incorporating a QTL (models B,C,E), the power with analyses 1 (that is, using the true map) was in the ranges 0·06–0·09 for $n=100$, 0·09–0·15 for $n=300$ and still only 0·33–0·38 for $n=1000$ (Table 2). The true significance level of these experiments is 0·05, by definition of the thresholds used.

Comparison of these results using the true map (cases B1,C1,E1) with other results for the same model but with incorrect maps used in the analysis portray the effect of incorrect maps on the power of QTL mapping. Again, differences between analyses using the true maps and those using incorrect maps were small. None of the results was significantly different from the corresponding result for Map 1 when tested using the normal approximation for a difference in binomial proportions (Table 2).

Estimated QTL positions for some of the analyses with all markers assumed to be on only one chromosome and with $n=1000$ are shown in Fig. 3 in order to illustrate features of the results. Unless stated otherwise, we refer to the position with the highest likelihood as being the estimated position of a QTL for a particular simulated population, whether or not the test statistic exceeded the 5% threshold. For case A1 (no QTL modelled), 47% of the simulated populations had an estimated QTL position at a marker position, with an additional 12% at the ends of the chromosome. Results were similar for other analyses of model A, with 41–52% of estimated QTL positions at a marker position and 10–13% at the ends of the chromosome. This trend was less evident for simulations with higher values of the test statistic. For example, in case A1, only 12 of the 50 simulations (24%) that exceeded the 5% threshold placed the QTL at a marker position, compared with 47% of all simulations for this case.

When a QTL was modelled at the third marker (model B), 20% of the simulations (whether exceeding the 5% threshold or not) using the correct map (case B1) positioned the QTL at that marker, whereas 51% (that is, an additional 31%) placed the QTL closer to that marker than to any other. Case B7 had 66% with the QTL positioned nearest the second and third markers (which were in the same position). The other cases (B2–B6,B8,B9) placed the QTL nearest the third marker in similar proportions (52–55%) to the analysis with the true map. The effect of an imprecise map that is expanded more on one side of the QTL than on the other is illustrated by case B6 (Fig. 3). In such cases, a moderate number of simulations place the QTL near the correct position on the more condensed side (in this case, between positions 20 cM and 25 cM), although there are similar numbers on each side (38% before the correct position, 39% after). The mean position is on the expanded side (at 32 cM), but this is due to positions on this side being stretched away from the true QTL position.

For the cases in which a QTL was modelled between a pair of markers (models C,E), we consider the proportion of times that the QTL (whether exceeding the 5% threshold or not) was positioned at or between those markers. Using the correct map, the QTL was positioned there in 52% and 40% of the simulations for cases C1 and E1, respectively. The use of incorrect maps gave similar figures (48–52% for cases C2, C3 and C9, and 39–41% for cases E2, E3 and E8) except for case C8. For this case, the QTL was positioned at or between the flanking markers 62% of the time, but this map included another marker incorrectly positioned in this interval. This figure was similar to the proportion (65%) of simulations in which the QTL was positioned at or between the first and third markers when using the correct map.

Mean translated distances to the true QTL position with $n=1000$ are shown in Table 3 for all simulations and for those that were significant at the 5% level.

https://doi.org/10.1017/S0016672304006974 Published online by Cambridge University Press

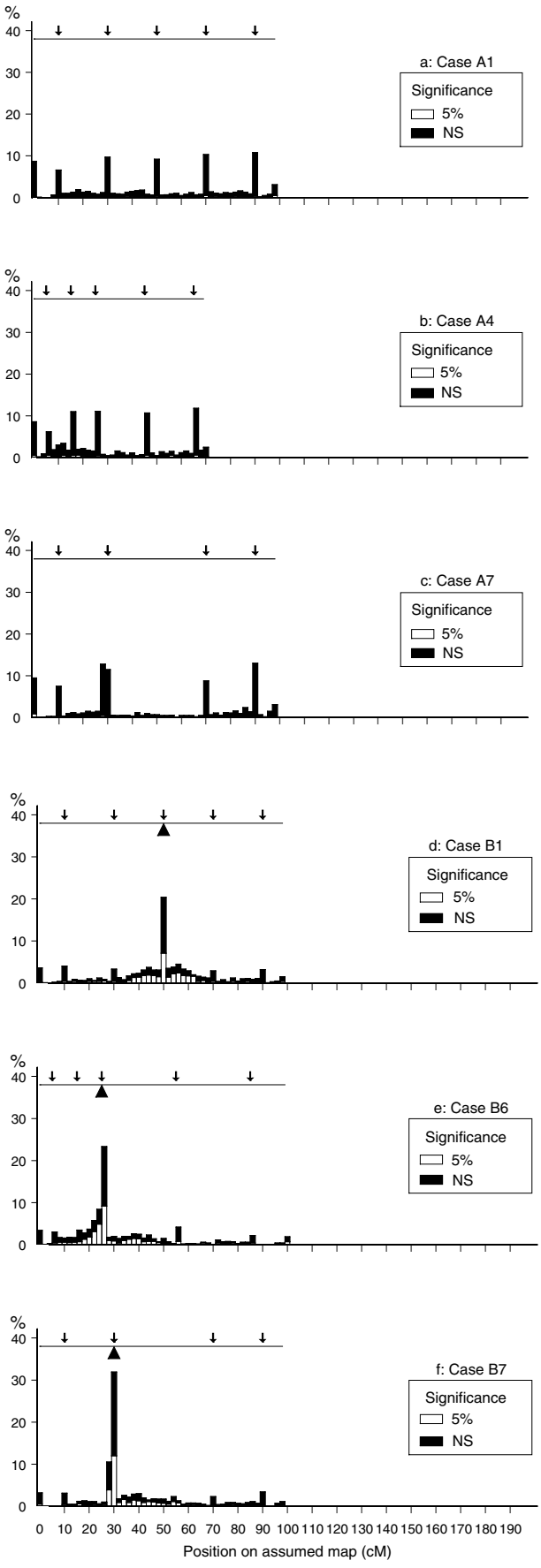

Fig. 3. (*Cont.*)

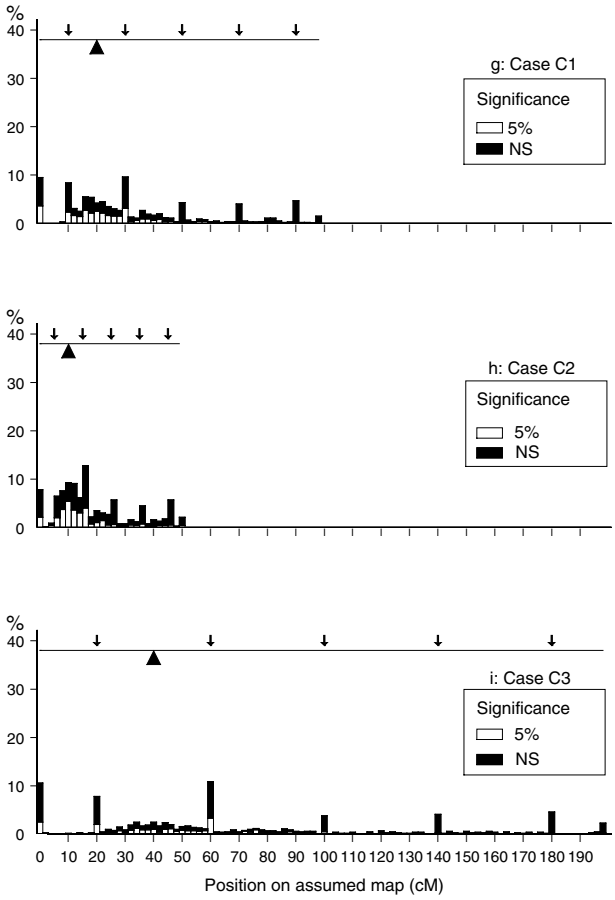

Fig. 3. Placement of putative QTLs for 1000 replicates of selected models and map combinations. Arrows and triangles (where relevant) show marker positions and the simulated QTL position, respectively. A horizontal line below the marker positions represents the chromosome(s) scanned. (a) Case A1 (i.e. model A, analysis 1). (b) Case A4. (c) Case A7. (d) Case B1. (e) Case B6. (f) Case B7. (g) Case C1. (h) Case C2. (i) Case C3.

For an unbiased method, these have expected value zero. When using the correct map, the means over all simulations were −0·9 cM, 13·4 cM and 14·8 cM for models B, C and E, respectively. Only the first of these is not significantly different from zero. Analyses using incorrect maps but with all markers on the correct chromosome did not give significantly different mean distances compared with the corresponding analyses with the true map, apart from case C2, in which the mean translated distance to the true QTL position was increased to 16·1 cM ($P < 0{\cdot}05$).

Estimated QTL positions for the analyses with one of the markers placed on a separate chromosome and with $n = 1000$, are shown in Fig. 4. Case A9 placed the QTL at the ends of chromosomes more often (16%) than case A1, with a corresponding reduction in placement at marker positions (40%). Case B9 placed the QTL at the correct position 23% of the time and closer to the correct marker than any other 52% of the time, similar to results with map 1 for this model.

https://doi.org/10.1017/S0016672304006974 Published online by Cambridge University Press

Table 3. *Mean translated distances (cM) between the estimated QTL position and the true QTL position for simulations with* n = *1000*

		All simulations		Significant simulations[b]	
Model[a]	Map	Mean distance to QTL	Standard deviation	Mean distance to QTL	Standard deviation
B	1	−0·9	21·7	0·3	15·4
	2	−0·9	21·4	−1·0	16·4
	3	−1·5	21·9	−1·0	16·2
	4	0·2	21·4	1·0	15·7
	5	−0·4	20·4	0·1	15·5
	6	−0·2	21·0	0·7	16·1
	7	−0·4	21·2	−0·7	16·4
	8	−2·5	19·5	−3·7	14·1
C	1	13·4	25·8	4·3	17·1
	2	16·1*	26·7	8·6	21·2
	3	15·3	27·4	8·7	19·7
	8	15·4	25·0	10·0	18·8
E	1	14·8	25·8	7·1	17·5
	2	13·7	25·4	5·8	17·9
	3	15·0	25·7	8·2	18·9
	8	13·4	25·3	6·6	16·2

[a] Refer to Table 1 for the characteristics of each model.
[b] Results for simulations where the QTL was significant at the 5% level.
* $P < 0{\cdot}05$ when comparing with analysis 1.

For case C9, the QTL was closer to one of its flanking markers than any other marker 67% of the time, compared with 70% for map 1 for this model.

For case A9, in which there was no QTL modelled and one marker was placed on a separate chromosome, the QTL was placed on the chromosome with the single marker 19% of the time, roughly in proportion to the number of markers (1 out of 5) on that chromosome. Similarly, other analyses for model A placed the QTL closest to each marker approximately 20% of the time, no matter what map was assumed. By contrast, analyses of model D simulations tended to place the QTL closest to the markers that were more widely spread (on the true map), no matter what map was assumed for the analysis.

4. Discussion

Knott and Haley (1992) found that an incorrectly scaled map had little effect on estimates of QTL position and effect size, and no influence on the test statistic for interval mapping in an F_2 population. Similarly, Hyne *et al.* (1995) found little difference in estimates of QTL position and effect size when the map was estimated from the QTL mapping population rather than when the true map was used. The present study extends these results to a wider range of incorrect maps. We have found that the accuracy of the map generally has little or no impact on the ability to detect QTLs, the true level of the test or the relative placement of QTLs. The number of times an imprecise map caused a significant difference ($P < 0{\cdot}05$) in the power or size of the test (once out of 78 comparisons) or in the estimated position (once out of 13 comparisons) were no higher than would be expected if there is no effect. In contrast to these results, Göring & Terwilliger (2000) found that treating the map as a nuisance parameter gave some increase in power. Their study involved a disease locus (rather than QTL mapping) with closely spaced but not fully informative markers. The increase in power might be due to these characteristics or to an alternative analysis method, rather than considering the effect of an imprecise map *per se*.

Even when one of the markers flanking the QTL is assigned to the wrong linkage group (case C9), there was only a minor drop in power (0·32 compared with 0·33 with the correct map; $P = 0{\cdot}37$). A drop in power similar to that from using single marker analyses rather than interval mapping (Darvasi *et al.*, 1993; Lander & Botstein, 1989; Rebai *et al.*, 1995) might have been expected, because the flanking marker information has been lost with this analysis. However, a simple comparison of the number of times the test statistic exceeds a set threshold does not take into account any possible change in significance level. Case A9 indicates that the significance level with this assumed map has increased from 0·05 to 0·06. If we ensure the same significance level (0·05) for analyses 9 as for other analyses by adjusting the critical value, we find the powers of cases B9 and C9 drop to 0·35 ($P = 0{\cdot}29$) and 0·29 ($P = 0{\cdot}03$), respectively, below that when using the correct maps. This latter significance test does not account for the process of estimating the

https://doi.org/10.1017/S0016672304006974 Published online by Cambridge University Press

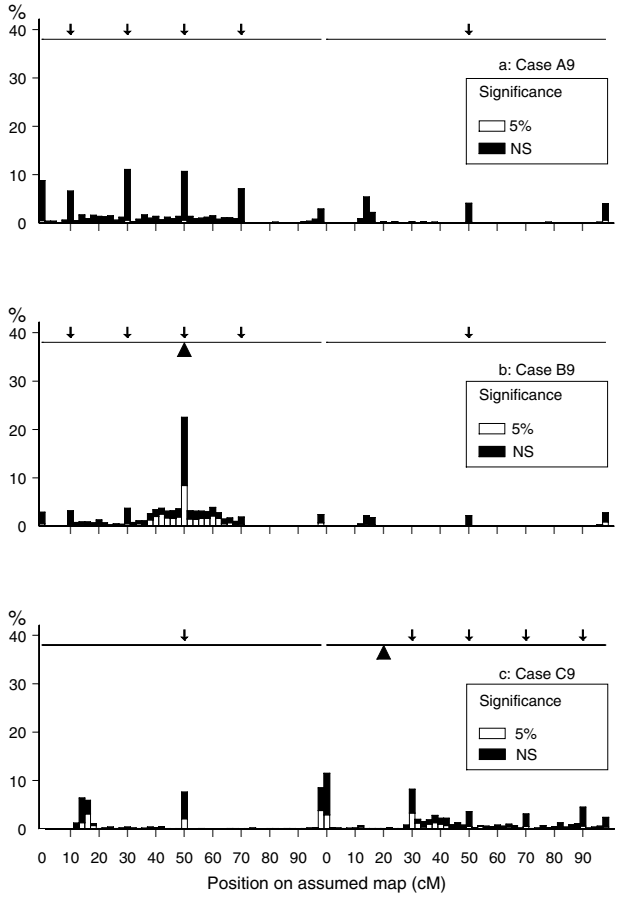

Fig. 4. Placement of putative QTLs from simulations with one marker placed on a separate linkage group. Arrows and triangles (where relevant) show marker positions and the simulated QTL position, respectively. Horizontal lines below the marker positions represent the chromosomes scanned. The chromosome with a single marker is placed on the side corresponding to its true position, but the orientation of that chromosome is arbitrary. (a) Case A9. (b) Case B9. (c) Case C9.

critical value, and so will be too liberal, but nevertheless suggests a slight drop in power. This drop in power (6–14%) is similar to the levels (3–7%) found in previous studies comparing interval mapping and single marker analyses with 20 cM marker spacing and intermediate levels of power (Darvasi *et al*., 1993; Rebai *et al*., 1995).

Lander & Botstein (1989) determined the threshold for a single 100 cM chromosome with 20 cM marker spacing to be 6·45. Using this threshold, we obtain type I error rates of 0·045 and 0·051 for cases A1 and D1, respectively, confirming this threshold. We have found that an incorrect map has little influence on the type I error rate (apart from cases in which a marker is assigned to another chromosome). The type I error, therefore, appears to be determined by the true map rather than the assumed map. Error rates determined by simulation methods (Churchill & Doerge, 1994) will therefore also be determined by the true map, so such methods will provide appropriate significance levels with an incorrect map even if an improved map becomes available after the QTL analysis. However, if analytical techniques are used to calculate an appropriate threshold (Lander & Botstein, 1989; Rebai *et al*., 1994), maps that are too sparse will give higher thresholds and maps that are too dense will give lower thresholds and so give incorrect type I error rates.

In our simulations, we have found a disproportionate placing of QTLs at markers to a degree that appears not to depend on the assumed map (Fig. 3). The procedure we used evaluated likelihoods at 2 cM intervals, so some of these maxima might have occurred close to, rather than at, marker positions. Darvasi *et al*. (1993) also observed that interval mapping tended to place a QTL at a marker, but did not note to what degree this occurred. Liu & Dekkers (1998) and Walling *et al*. (1998, 2002) also found this effect, particularly when the QTL variance was low.

Although case B7 had a higher proportion of simulations with the QTLs positioned correctly than the analysis with the correct map, this does not mean the analysis is actually improved. Map 7 incorrectly places the second and third marker in the same position, so this higher proportion is an artefact of the true QTL position representing 20 cM on the true map. A fairer comparison of these analyses is how often the QTL is placed at or between these markers (Fig. 3). These are 45% for Map 1 and 32% for Map 7. In another, fairer, comparison, 67% of the simulations with Map 1 placed the QTL nearer the second or third markers than any other marker, comparable to the 66% obtained with Map 7.

For case B6, we found that mean estimated QTL position is on the expanded side (at 32 cM), but this is due to positions on this side being stretched away from the true QTL position (Fig. 3). To see this, we look at the translated positions (Table 3) and find a mean that coincides with the true QTL position (within sampling error). Therefore, the bias in estimating the QTL position is accounted for by correction of the map, so that the bias introduced by analysis with the incorrect map can be corrected if the 'true' map becomes available. In general, we have found that incorrect maps have little influence on the translated positions when all markers are on the correct chromosome. One exception to this was for model C, with a uniformly condensed map. The bias in estimated QTL position, even with the correct map, for models C and E is due to there being more of the chromosome to the right than the left of the QTL (see below). However, this bias significantly increased when the analysis assumed a map that was uniformly condensed (Case C3). It is unclear why this occurred, particularly given that the uniformly expanded map (Map 3) also gave an increase in the bias, although

not significantly. The results might be due to the simulations for Map 1 giving a lower bias than they should have, and therefore lower bias than other analyses for this model (Table 3). This is supported by the fact that simulations at the other sample sizes (100 and 300) did not replicate this result.

The biases in translated positions for models C and E are due to the fact that the QTL was not centred on the chromosome for these models, and therefore there were more possible positions on one side of the QTL than the other. The bias is always on the side with more of the chromosome (using the true map). The bias decreased as the threshold increased, if we consider only those simulations where the threshold was exceeded. A similar effect was noted by Knott & Haley (1992), Hyne *et al*. (1995), Visscher *et al*. (1996), Knott *et al*. (1997), Scheler *et al*. (1998), Walling *et al*. (1998) and Charmet (2000) when the size of the QTL effect was small and/or the QTL was situated far from the centre of the chromosome. However, Knott & Haley (1992) did not find this effect for QTLs that explained at least 3% of the phenotypic variance and, in one such case, the estimated positions were biased in the opposite direction. This is probably due to the effect noted above (preferential placement of the QTL at a marker position), although this effect was not explicitly noted by Knott & Haley (1992). In their simulations with a non-centrally placed QTL, the closest marker was on the shorter end of the chromosome. Walling *et al*. (1998) show the combined effect of these two sources of bias by studying a range of QTL positions and marker spacings. We found the size of the bias for these models was reduced when considering results with higher values of the test statistic, as did some of the previously mentioned studies.

Although we found no particular problems or features of having a map with two markers incorrectly specified at the same position (analyses 7), this might not be the situation with other types of data and analysis. In particular, a recombinant between a pair of markers specified to be in the same position causes difficulties with the program Animap (Nielsen *et al*., 1995), which is used for outcrossing species in which individuals are not always informative at a particular marker. This is because a likelihood for genotypes and markers is calculated and, if there is recombination between markers in the same position, this likelihood will be zero regardless of the QTL position.

When one marker was placed on another chromosome, the distribution of estimated QTL positions was similar to the other cases in terms of how often they were placed closer to the closest marker(s) than any other marker. In these cases, the QTL was positioned at chromosome ends more often than for other analyses for the same model, but this reflects an increased opportunity to do so (four chromosome ends instead of two). For case C9, when the QTL was positioned on the extra chromosome with the single marker (36% of the time), it was often positioned far from the marker (for example, it was placed further than 20 cM from the marker 71% of the time). A similar effect was noted by Knott & Haley (1992), particularly when the QTL effect was small. The ability of QTL mapping to place a QTL relative to a single marker is poor (Weller, 1986) and so little importance would be attached to such distance estimates.

The placement of QTLs when there are none does not appear to be proportional to the assumed map distance. Rather, it appears to be more closely related to the distance spanned by a marker, because we found an approximately even distribution with respect to closeness to markers when the markers were evenly spaced (model A), but QTLs were more often placed closest to the more widely spread markers when they were unevenly spaced (model D).

A potential limitation of the present study is that actual map errors might differ from the types or in degree from those considered here. Alternative approaches and possible future work include: (1) simulation of map estimation for each iteration (similar to Hyne *et al*. (1995) but allowing for genotyping errors); and (2) Bayesian or profile likelihood methods (Göring & Terwilliger, 2000) to incorporate map uncertainty in the QTL analysis. We have, however, considered a wide range of possible types of map errors, to a degree that would be considered quite severe.

Our results have shown that the ability to detect QTLs, the true level of the test and the relative placement of QTL are mainly unaffected by the accuracy of the map. Therefore, QTL mappers do not need to expend a large effort in determining an appropriate map before proceeding with QTL analyses. QTL positions estimated using one map can be easily translated to the position that would have been obtained if an alternative (e.g. updated) map was used. Published comprehensive marker maps will normally be the most reliable source of linkage group membership and so some of the minor problems we have found with incorrectly assigning a marker to another linkage group can usually be avoided.

Funding for this project was provided by the New Zealand Foundation for Research, Science and Technology under contract CO4616.

References

Basten, C. J., Weir, B. S. & Zeng, Z.-B. (1994). ZMAP – a QTL cartographer. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production* **22**, 65–66.

Basten, C. J., Weir, B. S. & Zeng, Z.-B. (1997). *QTL Cartographer: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University.

Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–331.

Bovenhuis, H., van Arendonk, J. A. M., Davis, G., Elsen, J.-M., Haley, C. S., Hill, W. G., Baret, P. V., Hetzel, D. J. S. & Nicholas, F. W. (1997). Detection and mapping of quantitative trait loci in farm animals. *Livestock Production Science* **52**, 135–144.

Charmet, G. (2000). Power and accuracy of QTL detection: simulation studies of one-QTL models. *Agronomie* **20**, 309–323.

Churchill, G. A. & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Darvasi, A., Weinreb, A., Minke, V., Weller, J. I. & Soller, M. (1993). Detecting marker–QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**, 943–951.

Göring, H. H. H. & Terwilliger, J. D. (2000). Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *American Journal of Human Genetics* **66**, 1298–1309.

Hyne, V., Kearsey, M. J., Pike, D. J. & Snape, J. W. (1995). QTL analysis: unreliability and bias in estimation. *Molecular Breeding* **1**, 273–282.

Kearsey, M. J. & Farquhar, A. G. L. (1998). QTL analysis in plants; where are we now? *Heredity* **80**, 137–142.

Knott, S. A. & Haley, C. S. (1992). Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. *Genetical Research* **60**, 139–151.

Knott, S. A., Neale, D. B., Sewell, M. M. & Haley, C. S. (1997). Multiple marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. *Theoretical and Applied Genetics* **94**, 810–820.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Liu, Z. & Dekkers, J. C. M. (1998). Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. *Genetics* **148**, 495–505.

Nielsen, D., Cockett, N. E. & Georges, M. (1995). Mapping markers and quantitative traits in large half-sib pedigrees. *Proceedings of the Western Section of the American Society of Animal Science* **46**, 205–208.

Rebai, A., Goffinet, B. & Mangin, B. (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**, 235–240.

Rebai, A., Goffinet, B. & Mangin, B. (1995). Comparing power of different methods for QTL detection. *Biometrics* **51**, 87–99.

Scheler, P., Mangin, B., Goffinet, B., Le Roy, P., Boichard, D. & Elsen, J. M. (1998). Properties of a Bayesian approach to detect QTL compared to the flanking markers regression method. *Journal of Animal Breeding and Genetics* **115**, 87–95.

Tomfohrde, J., Wood, S., Schertzer, M., Wagner, M. J., Wells, D. E., Parrish, J., Sadler, L. A., Blanton, S. H., Daiger, S. P., Wang, Z. Y., Wilkie, P. J. & Weber, J. L. (1992). Human chromosome-8 linkage map based on short tandem repeat polymorphisms – effect of genotyping errors. *Genomics* **14**, 144–152.

Visscher, P. M., Haley, C. S. & Knott, S. A. (1996). Mapping QTLs for binary traits in backcross and F_2 populations. *Genetical Research* **68**, 55–63.

Walling, G. A., Visscher, P. M. & Haley, C. S. (1998). A comparison of bootstrap methods to construct confidence intervals in QTL mapping. *Genetical Research* **71**, 171–180.

Walling, G. A., Haley, C. S., Perez-Enciso, M., Thompson, R. & Visscher, P. M. (2002). On the mapping of quantitative trait loci at marker and non-marker locations. *Genetical Research* **79**, 97–106.

Weller, J. I. (1986). Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* **42**, 627–640.