Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

DETECTION AND UTILISATION OF ASSOCIATIONS BETWEEN GENETIC MARKERS AND QUANTITATIVE TRAIT LOCI IN RADIATA PINE

A thesis presented in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy in Forest Genetics and Breeding
at
Massey University
Palmerston North
New Zealand

Satish Kumar

2000

ABSTRACT

This thesis focuses on the detection of quantitative trait loci (QTL) and utilisation of marker-trait associations in tree breeding programs. Theoretical expressions of variance components for single-marker ANOVA were derived and were used in deterministic simulations to determine the power of two-generation QTL mapping experiments in an outbred population containing full-sib or self families. Analysis of one linkage group for mapping QTL for wood density is presented. Genetic markers that are linked to the QTL can be used for selection purposes. The effect of using genetic markers to assist in different selection and deployment schemes was studied. The additional response to selection was computed using stochastic simulation. Most of the studies on MAS are concentrated on expected genetic progress for a single trait. However, in practice, the commercial breeding objective is nearly always composed of several traits. For a conventional breeding framework, a method is proposed to reduce the effect of sampling errors on the estimates of multivariate genetic parameters and thus increasing the efficiency of index selection. The general discussion of this thesis addresses QTL analysis methods, various aspects of QTL mapping designs and implementation of MAS in the radiata pine breeding program in New Zealand.

Dedicated to Mum, Dad and Sandhya

ACKNOWLEDGEMENTS

I am grateful to Prof. Dorian Garrick, my chief supervisor, for giving me the opportunity to study under his supervision. I wish to express my great appreciation for his guidance, constructive criticism and encouragement during the completion of this thesis. I am also grateful to Dr Sue Carson, my co-supervisor, for her interest in the development of my PhD programme. I have benefited immensely from her ideas and support throughout.

I am thankful for Dr Rowland Burdon for his invaluable input in this thesis. Your promptness of reviewing and providing constructive comments on my manuscripts is highly appreciated.

I am highly appreciative of the inputs of scientists at *Forest Research* (Dr Phillip Wilcox, Dr Mike Carson, Dr Rod Ball, Dr Tom Richardson, Dr Tony Shelbourne and Dr Luis Gea).

Thanks to Dr Richard Spelman for many helpful discussions over the phone.

I am very much appreciative of the opportunity and the financial support given by *Forest Research* to undertake my PhD.

I wish to express my gratitude to all friends, especially Alastair Currie, Luis Apiolaza, Lisa Watson and Nicolas Lopez-Villalobos, at Massey University for making these four years as one of my wonderful experiences.

Thanks to the Director (Prof. Hugh Blair) and Administrator (Mrs. Allain Scott) of Research and Post-Graduate Studies for being helpful throughout.

I thank my parents, brothers and sister for their love and encouragement. Special thanks to my wife Sandhya for her love, patience and understanding.

CONTENTS

Chapter 1

GENERAL INTRODUCTION

Conventional breeding methods in radiata pine in New Zealand have been under way for almost fifty years and have lead to substantial genetic improvement. Long generation intervals typical of forest tree species necessitate efficient approaches to improvement programs. Molecular markers arguably have greater potential impact in tree improvement and breeding than in short-rotation crops because of their capacity to: minimise the general interval, increase the genetic gain per generation, and enable improvements to be rapidly deployed (DALE & CHAPARRO, 1996).

Most applications of genetic markers requires the mapping of quantitative trait loci (QTL), i.e., the dissection of genetic variance into components due to individual QTL. SAX (1923) was the first to show how genetic factors influencing quantitative traits can be identified by using molecular markers. In beans, Sax scored morphological traits with monogenic inheritance and found the seed weight of certain morphological variants to be significantly higher than the seed weight of other variants. He concluded that a size factor, which we would now call a QTL, was linked to one of the morphological markers studied. As a result of genetic linkage between the marker and the QTL, the size factor cosegregated with the genes underlying the morphological traits.

NEIMANN-SØRENSEN & ROBERTSON (1961), using blood groups in dairy cattle, showed how associations between markers and quantitative traits can be studied in outbred populations. They found no significant associations, which they attributed to the fact that they had only a few markers available which gave them a low chance of having a marker close to a QTL. With the advent of molecular technology, many genetic markers have become available. The construction of linkage maps with genetic markers covering the whole genome allows for systematic screening for genes or chromosomal regions influencing important traits (BOTSTEIN *et al.*, 1980). When marker-QTL associations have been identified and located to chromosomal segments, the marked QTL can be utilised in breeding schemes by marker assisted selection (MAS). Favourable theoretical genetic and economic responses to MAS have been reported (KERR *et al.*, 1996).

Aim

The aim of this thesis is to contribute to the efficient utilisation of genetic markers in tree breeding. This thesis studies the design of QTL mapping experiments, the issues of QTL detection and the use of marker assisted selection in an outbred population.

Outline of this thesis

The designs of linkage mapping experiments that involve crosses between inbred lines are well documented. For outbred populations, like forest trees, less research has been aimed at specifying the appropriate designs. From population genetic studies, it is known that wild allogamous species like forest trees are likely to be in linkage equilibrium (reviewed in STRAUSS *et al.*, 1992). Full-sib families are commonly used in QTL mapping studies in forest trees. For an outbred full-sib radiata pine pedigree, as many as four alleles may be segregating at a marker locus and male and female parents might be segregating for different QTL alleles. Therefore, this thesis studies designs for finding linkage between a marker locus and a trait locus in outbred populations. Chapter 2 of this thesis derives formulae for evaluating the power of linkage detection in full-sib families of outbred populations.

Self-families of outbred trees are also being used for studying the linkage between markers and the genes causing inbreeding depression in growth traits (FU & RITLAND, 1994). On the assumption that there will be some individuals without appreciable genetic load, the self-families have been used to detect growth-related QTL that would be expressed under the normal course of outbreeding (PLOMION *et al.*, 1996). Chapter 3 of this thesis provides some theoretical expressions for evaluating the power of QTL mapping designs for linkage detection in self-families of outbred populations.

Forest Research, New Zealand, established a QTL mapping experiment in a full-sib family of radiata pine. The objective of the experiment was to identify chromosomal regions that affect wood density and growth traits. In Chapter 4, linkage group three was analysed for associations between 16 RAPD markers and wood density measured at three different ages. Estimates for QTL effects and locations along with their bootstrap confidence intervals were obtained.

Once genetic markers that explain a significant amount of genetic variance have been identified, these markers can be used in breeding programs. Previous simulation studies have evaluated either across-family MAS or have applied MAS to an unselected population. Assuming that the linkage state of the marker alleles with respect to the QTL might vary between families, in Chapter 5 of this thesis, within-family MAS was applied to a population where prior selection had been undertaken for one to two generations.

Index selection is commonly applied in tree breeding programs. The efficiency of traditional index selection can be improved by selection on the QTL that have positive effects on the traits included in the index. However, multitrait MAS does not seem to be a practical option in the near future. Chapter 6 of this thesis considers aspects of improving the efficiency of index selection within a conventional breeding framework. Impacts of sampling errors on estimates of multivariate genetic parameters were evaluated and a new method to improve the efficiency of index selection is compared with some of the available alternatives for this purpose.

In the general discussion, firstly, the methods for detecting linkage are discussed. Secondly, the setting of critical values is discussed. Further, the different QTL mapping design options, which can be used for increasing the power of linkage detection, are discussed. Finally some thoughts on the possible applications of MAS in forest tree breeding programs in New Zealand are given.

REFERENCES

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.

DALE, G. & CHAPARRO, J. 1996: Integration of molecular markers into tree breeding and improvement programs. Pp. 472-477 *in*: M. J. DIETERS, A. C. MATHESON, D. G. NIKLES, HARWOOD, C. E. & MALKER, S. M. (eds). *Tree Improvement for Sustainable Tropical Forestry*. Proc. QFRI-IUFRO Conference, Caloundra, Queensland, Australia, 27 Oct – 1 Nov 1996.

FU, Y. B. & RITLAND, K. 1994: Marker based inferences about fecundity genes contributing to inbreeding depression in *Mimulus guttatus*. *Genome* **37**:1005-1010.

KERR, R. J., JARVIS, S. R. & GODDARD, M. E. 1996: The use of genetic markers in tree breeding programs. Pp. 498-505 *in*: M. J. DIETERS, A. C. MATHESON, D. G. NIKLES, HARWOOD, C. E. & MALKER, S. M. (eds). *Tree Improvement for Sustainable Tropical Forestry*. Proc. QFRI-IUFRO Conference, Caloundra, Queensland, Australia, 27 Oct – 1 Nov 1996.

NEIMANN-SORENSON, A. & ROBERTSON, A. 1961: The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta. Agric. Scan.* **11**:163-196.

PLOMION, C., DUREL, C.-E. & O'MALLEY, D. M. 1996: Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. *Theoretical and Applied Genetics* **93**: 849-858.

SAX, K. 1923: The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**:552-560.

STRAUSS, S. H., LANDE, R. & NAMKOONG, G. 1992: Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research* **22**:1050-1061.

Chapter 2

DETECTING LINKAGE BETWEEN A FULLY-INFORMATIVE MARKER LOCUS AND A TRAIT LOCUS IN OUTBRED POPULATIONS USING ANALYSIS OF VARIANCE

S. Kumar, S. D. Carson & D. J. Garrick

FOREST GENETICS (2000) 7: 47-56

ABSTRACT

Analysis of variance can be used to detect the linkage of segregating quantitative trait loci (QTL) to molecular markers in outbred populations. Given a single fully-informative (FI) marker for independent full-sib families (with marker configuration: M_1M_2 x M_3M_4) and assuming linkage equilibrium, variance components were derived to predict the power of detection of a QTL. These variance components are based on a hierarchical analysis of variance assuming a completely random model. Formulae that relate power to the recombination frequency (r) between the FI marker and the QTL, genetical properties of the quantitative trait controlled by the QTL and the design parameters are developed. The predicted powers using the FI marker configuration were compared to that obtained using pseudo-backcross (PBC : M_1M_2 x M_1M_1) and pseudo-intercross (PIC : M_1M_2 x M_1M_2) marker configurations. The effects of dominance properties of the QTL on power were also examined. The reliability of the theoretical approximation of power was confirmed by computer simulations. The results showed that: FI marker design is more efficient than PBC and PIC marker designs; few large families are better than many small families. Incomplete linkage and dominance of the QTL showed large effects on the power.

Key words: fully-informative marker, genetic linkage, statistical power, QTL

INTRODUCTION

The use of molecular markers as a complementary tool for breeding is based on linkage disequilibrium between marker and quantitative trait loci (QTL) involved in the control of quantitative characters. In most agricultural crops, inbreeding is followed by crossing between inbred lines to create disequilibrium for QTL detection. From population genetic studies, it is known that wild allogamous species like forest trees are often in linkage equilibrium and, because of the long generation intervals and inbreeding depression, it is difficult to obtain inbred lines for QTL mapping experiments. Linkage disequilibrium between a marker and a linked QTL, however, can be found within families in outcross populations and it is increasingly common to carry out QTL detection in a full or half-sib family.

In an outbred population different marker alleles will likely be associated with the same QTL allele in different families. Therefore, evidence for a linked QTL cannot be obtained at a population level from overall mean differences between marker genotypes. Using an hierarchical ANOVA, marker effects need to be analysed in each family separately and the test for a linked QTL comes from the comparison of the between-marker genotypes within-family mean squares with the residual mean squares and can be tested as an F-ratio (HILL 1975, SOLLER & GENIZI 1978). Under the null hypothesis (marker is not linked to the QTL i.e., recombination (r) between marker and QTL is 0.5), this ratio is distributed as a central F-variable; whereas this ratio will be a noncentral F-variable when r is less than 0.5 (JAYAKAR 1970, LUO 1993). Hence, given the pedigree structure, it is possible to predict the power of detection of a given QTL (HILL 1975, SOLLER & GENIZI 1978, LUO 1993, KNOTT 1994).

The informative full-sib families considered in previous simulation studies of outbred populations (HILL 1975, SOLLER & GENIZI 1978, LUO 1993, KNOTT 1994) were of two types with respect to the marker genotypes of the parents. First, those where one parent is homozygous at the marker locus and one is heterozygous (pseudo-backcross or PBC families) and second, those where both parents are heterozygous for the same genotype at the marker locus (pseudo-intercross or PIC families). Both of these strategies suggest that only two alleles are segregating in full-sib progeny. However, for an outbred *Pinus radiata* pedigree, as many as four alleles may be segregating at a locus. With the continued development of multiallelic codominant markers (for example, microsatellites), the exclusive use of fully-informative markers (i.e., $M_1M_2 \times M_3M_4$) is becoming possible. This creates an additional family type (fully-informative or FI families) with respect to the markers; that is, one where all four marker genotypes can be distinguished in the offspring.

HILL (1975) and LUO (1989, 1993), assuming a bi-allelic marker and the linked bi-allelic QTL, derived the expressions for expected variances for ANOVA of PBC and PIC marker designs in a segregating population. MURANTY (1996) derived the expressions for the noncentrality parameter of different mating schemes assuming FI marker design. However, these expressions could not relate the power directly to different recombination rates and non-additive gene action at QTL. So far, the

theoretical expressions of variance components, to predict the power of using FI marker type families using ANOVA, have not been derived.

The present study was focused on deriving the expressions of expected variances for ANOVA of a FI marker design ($M_1M_2 \times M_3M_4$), in two-generation pedigrees of outbred populations, and relating the power directly to genetic parameters at the QTL and the relevant design parameters. This will allow factors affecting the power to be investigated comprehensively. A second objective of this study was to compare the power obtained from using FI marker design to that obtained from PBC and PIC type marker strategies.

THEORY

Basic assumptions and experimental design

The underlying assumptions of the method are those commonly made by researchers. The method involves analysing progeny from controlled mating in a population. Two autosomal loci are considered, one of them affects a quantitative trait (QTL) while the other is a fully-informative marker. The two loci are linked with a recombination frequency of r ($s = 1 - r$). Let the frequency of allele Q_1 at the QTL be denoted as p ($q = 1 - p$), and the phenotypic distributions of the 3 genotypes at the QTL i.e., Q_1Q_1, Q_1Q_2 and Q_2Q_2 are assumed to be $N(a, \sigma^2)$, $N(d, \sigma^2)$ and $N(-a, \sigma^2)$ respectively, where a and d represent the additive and dominance effects at the QTL. With just one QTL and no other gene effects, σ^2 will be only the environmental variance, whereas in the presence of unlinked QTLs, it will also include genetic variance at these loci (i.e., polygenic variance).

Let the parental genotypes at the marker locus be M_1M_2 and M_3M_4 and four marker genotype classes ($M = 4$) are distinguishable in the offspring: M_1M_3, M_1M_4, M_2M_3 and M_2M_4 segregating with a 1:1:1:1 ratio. We assume that the QTL and the marker gene are in linkage equilibrium in the population. Let n_{ij} denote the number of sibs within the j^{th} marker genotype class within the i^{th} sibship. Also each sibship (N_f) has a constant size of N_o and thus the total experimental size is $N_f \times N_o$.

Statistical Model

The linear model for the phenotype of the quantitative trait measured on the k^{th} sib (k = 1, 2, . . ., n_{ij}) with the j^{th} marker genotype (j = 1, 2, . . ., M=4) within the i^{th} sibship (i = 1, 2, . . ., N_f) can be written as:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk} \quad (1)$$

where μ is an overall mean, α_i, β_{ij} and e_{ijk} are contributions from the sibship, from the marker genotype within sibship and from within-marker within-sibship residual, respectively. They are assumed to be independently and normally distributed with zero means and variances σ_α^2, σ_β^2 and σ_e^2, respectively. Similar assumptions have been made in several studies (e.g., HILL 1975, LUO 1993, LYNCH & WALSH 1997). The assumption of functional independence of the quantitative trait from the marker locus was also made. We have considered only one QTL and all other " background" genetic variation is considered as environmental. The ANOVA for this model is given in Table 1.

Table 1. ANOVA for a two-factor completely nested design.

Source	Degrees of freedom	MS	EMS
Between sibships	N_f -1	MS_s	-
Between marker genotypes within sibship	$\sum(M_i - 1)$	MS_m	$\sigma_e^2 + n_0\sigma_\beta^2$
Within marker genotype within sibship	$\sum(n_{ij} - 1)$	MS_w	σ_e^2

Under the assumption of a constant size of sibship (N_o) and

$$n_{i1} : n_{i2} : n_{i3} : n_{i4} \approx 1 : 1 : 1 : 1, \quad (2)$$

the approximation for n_0 will be : $n_0 \approx 0.25\,N_o$. (3)

The expression for mean squares and the general version of n_0 (Table 1) can be found, for example, in HILL (1975). All possible marker-QTL genotypes of parents and the gametes inherited by the offspring are given in Table 2 with their probabilities. Using

these probabilities, the expected values of the quantitative trait value, y, were obtained for different marker genotypes within sibship (Appendix 1). Similarly, the variances of the trait value within sibship within marker genotypes were derived (Appendix 2). Finally, the variance between marker genotype classes within sibship were obtained (Appendix 3).

Table 2. Probabilities of various gametes inherited from parents to progeny. Parental marker genotypes are: M_1M_2 x M_3M_4. We assumed that the QTL and the marker genes are in linkage equilibrium in the population. Recombination rate between the marker and the QTL is r.

From first parent				
Parental genotype	Gametes			
	M_1Q_1	M_1Q_2	M_2Q_1	M_2Q_2
M_1Q_1 / M_2Q_1	1/2	0	1/2	0
M_1Q_1 / M_2Q_2	(1 - r) / 2	r/2	r/2	(1 - r)/2
M_1Q_2 / M_2Q_1	r/2	(1 - r)/2	(1 - r)/2	r/2
M_1Q_2 / M_2Q_2	0	1/2	0	1/2
From second parent				
Parental genotype	Gametes			
	M_3Q_1	M_3Q_2	M_4Q_1	M_4Q_2
M_3Q_1 / M_4Q_1	1/2	0	1/2	0
M_3Q_1 / M_4Q_2	(1 - r) / 2	r/2	r/2	(1 - r)/2
M_3Q_2 / M_4Q_1	r/2	(1 - r)/2	(1 - r)/2	r/2
M_3Q_2 / M_4Q_2	0	1/2	0	1/2

The variance expressions, given in Appendix 3, were averaged by using the corresponding probabilities as weights and it gives us the expected variance between marker genotypes within sibships (σ^2_β) as:

$$\sigma^2_\beta = pq\{(s-r)^2[(d+a)^2 - 2pqd^2 - 4pad] + 2pqd^2[(s^2+r^2)^2 + 4r^2s^2 - 0.5]\}. \quad (4)$$

Using Appendix 2, first the average variance within each marker genotype was obtained by using the corresponding probabilities as weights. After this, the variances within each marker genotype were averaged using equal probabilities (because four marker genotypes are assumed to be segregating with a 1:1:1:1 ratio) and it gives us the expected variance within marker genotypes within sibships (σ_e^2) as:

$$\sigma_e^2 = \sigma^2 + 4pqrs[p^2(a-d)^2 + q^2(a+d)^2 + 2pqa^2 + pqd^2(1-2rs)] + 2p^2q^2d^2(r^2+s^2)[1-(r^2+s^2)]. \quad (5)$$

The rationale used for the derivation of (4) and (5) is similar to that of HILL (1975). From equation (4) it can be easily shown that the expected variance between marker genotypes within sibship (σ_β^2) will be zero if there is no linkage between the marker and the QTL, i.e., $r = s = 0.5$. Under the null hypothesis ($H_0 : r = 0.5$) the ratio MS_m / MS_w has an expected value of 1 and is distributed as a central F-variable; whereas this ratio has an expected value of more than one and will be a noncentral F-variable when r is less than 0.5 (JAYAKAR 1970, LUO 1993). Using the standard definition, the power function for linkage detection with the design under study (FI families) can be written in the following general form:

$$\text{Power} = Pr\,[F_{(\nu 1,\, \nu 2;\, \delta)} > F_{(\alpha,\, \nu 1,\, \nu 2)}]\,, \quad (6)$$

where $F_{(\nu 1,\, \nu 2;\, \delta)}$ is a noncentral F-variable with degrees of freedom ν_1 and ν_2 and noncentrality parameter δ, while $F_{(\alpha,\, \nu 1,\, \nu 2)}$ is the upper α point of a central F-variable with degrees of freedom ν_1 and ν_2. The value of noncentrality parameter, δ, was calculated as (LUO 1993):

$$\delta = (M \times N_f - 1)\, n_0\, \sigma_\beta^2 / \sigma_e^2. \quad (7)$$

Power calculation

The power of a test is defined as the probability of rejecting the null hypothesis when its alternative is true. The power of a QTL mapping experiment is the probability that the null hypothesis (no linked QTL) is rejected when its alternative (presence of a linked

QTL) is true. The formulae developed for expected variance between marker genotypes within sibships (σ_β^2) and the expected variance within marker genotypes within sibships (σ_e^2) were used in theoretical prediction of the powers of QTL detection for a wide range of combinations of parameters (i.e., genetic parameters at the QTL and design parameters). In order to derive the parameters, the total genetic variance, V_G (sum of additive, V_A, and dominance, V_D, variance), arising from one locus (QTL) can be written as (FALCONER 1989):

$$\begin{aligned} V_G &= V_A + V_D \\ &= 2pq[a + d(q - p)]^2 + [2pqd]^2 \\ &= 2pq[a^2 + (1 - 2pq)d^2 + 2(q - p)ad]. \end{aligned} \quad (8)$$

By assuming the phenotypic variance (V_P) to be unity, the V_G (or QTL variance) becomes the broad-sense heritability (H^2) at the QTL. Also, $\sigma^2 = 1 - V_G$. To determine the value of parameter a and d at the QTL, we take following steps (LUO 1993):

Assume the dominance ratio (f) = d/a, then

$$a = \sqrt{\frac{V_G}{2pq[1+(1-2pq)f^2+2(q-p)f]}}, \text{ and} \quad (9)$$

$$d = f \times a. \quad (10)$$

Using different combinations of design parameters (N_f and N_o), genetic parameters at the QTL (p, f, and H^2) and recombination frequency (r), the noncentrality parameter can be calculated. After that power can be easily calculated using (6).

Power evaluation from simulations

Since approximations (2) and (3) were made in deriving the power function, the reliability of these approximations was checked by comparing the theoretical predictions of the power to the powers calculated from simulation experiments. A program was written in SAS (1989) for simulating the inheritance of marker-QTL linkage for any combination of experimental design and genetic parameters. The simulated data was analysed using SAS PROC GLM and the frequency of significant F-values in replicated simulation trials was calculated as in CARBONELL *et al.* (1992) and LUO (1993), which gives the empirical power.

Comparison of power

The power, calculated using FI families ($M_1M_2 \times M_3M_4$) in this study, were compared to those obtained from using the families where parents are $M_1M_1 \times M_1M_2$ (PBC families) or $M_1M_2 \times M_1M_2$ (PIC families). The power for these two designs (PBC and PIC) in a segregating population were evaluated by LUO (1993). The results for the PBC and PIC type marker configurations in our study are solely based on the formulae derived by LUO (1993).

RESULTS

Theoretical powers of linkage detection were calculated for a wide range of genetic parameters at the QTL and design parameters. Empirical powers, based on 500 replications, are presented along with those obtained from theoretical approximation. When assuming gene action at the QTL to be purely additive, the power of QTL detection for three types of marker loci varies substantially (Table 3).

Table 3. Theoretical prediction (PR) of powers of 3 marker designs for a QTL that has a heterozygosity of 50%, for various number of families (N_f), various number of offspring per family (N_o). The other assumptions were: broad-sense heritability at the QTL (H^2) = 0.05, recombination rate (r) = 0.10, type-I error = 0.01 and dominance ratio (f) = 0.0. The powers evaluated from simulation experiments (SI) are also given. PBC = pseudo-backcross, PIC = pseudo-intercross, FI = fully-informative.

		PBC		PIC		FI	
N_f	N_o	PR	SI	PR	SI	PR	SI
5	50	0.04	0.05	0.03	0.04	0.05	0.05
	100	0.11	0.12	0.07	0.08	0.15	0.15
	200	0.29	0.32	0.18	0.24	0.46	0.48
10	50	0.07	0.07	0.04	0.05	0.09	0.07
	100	0.21	0.21	0.12	0.12	0.30	0.28
	200	0.57	0.55	0.36	0.39	0.79	0.79
20	50	0.13	0.12	0.07	0.08	0.17	0.17
	100	0.41	0.41	0.22	0.22	0.58	0.57
	200	0.87	0.83	0.66	0.67	0.98	0.98

The power is the highest with FI markers (both parents have different heterozygote genotypes at marker locus) and is lowest for PIC markers (parents are heterozygous for the same genotype at marker locus). The power of linkage detection increases as the number of offspring per family increase. Keeping the number of offspring genotyped fixed (say, 1000), then having fewer larger families clearly increases power relative to many small families.

The various levels of genetic variance or the broad-sense heritability at the QTL (H^2) and different recombination rates between marker and the QTL has significant impact on power of QTL detection for all three marker configurations (Table 4). As the heritability at the QTL increase the power also increases but a decreasing trend in power was obtained for an increase in recombination rate. For a larger $H^2 = 0.15$, the theoretical powers of linkage detection when $r = 0.10$ were 0.68, 0.86 and 0.98 for PIC, PBC and FI marker loci, respectively, for a sample size of 10 families with 100 offspring each (Table 4). It also shows that once r is greater than 0.10 the power of linkage detection is very low even if the broad-sense heritability at the QTL is 0.15.

Powers were also evaluated with varying dominance ratio at the QTL (Table 5). It shows that the theoretical power of linkage detection increases, in general, for PIC and FI marker designs whereas it remains constant for PBC designs as the dominance increases. However, the rate of increase is large when small number of families with large number of offspring are used. For example the power increases from 0.18 to 0.24 and 0.46 to 0.53 for PIC and FI marker designs with a sample size of 5 families each having 200 offspring. The effect of different QTL allele frequencies on the power of linkage detection is shown in Table 6. For additive gene action at the QTL, power of linkage detection is highest when $p = 0.50$.

Table 4. Comparison of theoretically predicted (PR) powers of linkage detection of 3 marker designs for varying number of families (N_f) and number of offspring per family (N_o) where H^2 and r represent the broad-sense heritability at the QTL and recombination frequency between marker and the QTL. The other assumptions were: type-I error = 0.01, dominance ratio (f) = 0.0 and p = 0.50. The powers evaluated from simulation experiments (SI) are also given. PBC = pseudo-backcross, PIC = pseudo-intercross, FI = fully-informative.

		$N_f = 10, N_o = 100$						$N_f = 25, N_o = 40$					
		PBC		PIC		FI		PBC		PIC		FI	
H^2	r	PR	SI	PR	SI	PR	SI	PR	SI	PR	SI	PR	SI
0.05	0.0	0.41	0.43	0.24	0.27	0.60	0.58	0.23	0.27	0.12	0.11	0.32	0.30
	0.1	0.21	0.21	0.12	0.11	0.30	0.28	0.11	0.12	0.06	0.06	0.14	0.14
	0.3	0.03	0.03	0.02	0.01	0.04	0.03	0.02	0.04	0.02	0.02	0.02	0.02
0.10	0.0	0.87	0.87	0.69	0.67	0.98	0.97	0.68	0.65	0.42	0.46	0.86	0.87
	0.1	0.59	0.56	0.38	0.39	0.81	0.77	0.36	0.37	0.19	0.18	0.51	0.51
	0.3	0.08	0.08	0.04	0.05	0.10	0.06	0.04	0.05	0.03	0.03	0.05	0.04
0.15	0.0	0.99	0.96	0.94	0.86	0.99	0.99	0.94	0.90	0.76	0.76	0.99	0.99
	0.1	0.86	0.88	0.68	0.67	0.98	0.96	0.67	0.66	0.41	0.40	0.85	0.84
	0.3	0.14	0.15	0.08	0.09	0.20	0.21	0.08	0.10	0.04	0.06	0.09	0.09

Table 5. Effect of gene action at the QTL on power of linkage detection of 3 marker designs. The symbols f (= d/a) denotes the dominance ratio at the QTL. The powers given here were evaluated from theoretical prediction (PR) and simulation (SI) at H^2 = 0.05, r = 0.10, p = 0.50 and type-1 error = 0.01.

		PBC		PIC		FI	
N_f (N_o)	f	PR	SI	PR	SI	PR	SI
5 (200)	0.0	0.29	0.32	0.18	0.24	0.46	0.48
	0.5	0.29	0.35	0.20	0.26	0.49	0.49
	1.0	0.29	0.36	0.24	0.31	0.53	0.60
10 (100)	0.0	0.21	0.21	0.12	0.12	0.30	0.28
	0.5	0.21	0.23	0.13	0.15	0.32	0.29
	1.0	0.21	0.24	0.16	0.18	0.36	0.38
20 (50)	0.0	0.13	0.12	0.07	0.08	0.17	0.17
	0.5	0.13	0.15	0.08	0.08	0.18	0.17
	1.0	0.13	0.15	0.09	0.11	0.20	0.22

Table 6. Effect of gene frequency (p) at the QTL on power of linkage detection of 3 marker designs. The powers given here were evaluated from theoretical prediction at r (recombination rate) = 0.00, f (= d/a) = 0 and type-1 error = 0.01. Half the difference between QTL homozygotes (i.e., a) = 0.30 SD.

N_f (N_o)	p	PIC	PBC	FI
5 (200)	.20	.15	.25	.39
	.40	.29	.44	.67
	.50	.31	.46	.70
	.60	.29	.44	.67
	.80	.15	.25	.39
10 (100)	.20	.10	.17	.25
	.40	.19	.33	.49
	.50	.20	.35	.52
	.60	.19	.33	.49
	.80	.10	.17	.25

DISCUSSION

Power of QTL detection in two-generation outbred pedigrees with varying dominance ratios, size of the QTL, recombination rates between the marker and the QTL and design parameters, was predicted in the present study. Three types of marker configurations were investigated. Derivations in the present paper have shown that the power of detecting linkage between a fully-informative marker and a QTL can be expressed as function of design parameters and parameters describing genetic properties of the QTL. A very close agreement was found between the powers from theoretical evaluation and stochastic simulation under wide range of situations.

Effect of family- type on power

Three types of informative marker configurations in full-sib families were considered in the present study. First, those families where one parent is homozygous for marker ($M_1M_1 \times M_1M_2$, backcross-type or PBC family); second, those where both parents are heterozygous, with the same genotype at marker locus ($M_1M_2 \times M_1M_2$, intercross-type or PIC family); and third, those where both parents have different genotypes at marker locus ($M_1M_2 \times M_3M_4$, fully-informative or FI family). The power of the third-type of maker configuration was clearly the highest compared to the other two designs for all parameter combinations considered in this study (Table 3). This is because the use of a fully-informative marker allows all four genotypic classes to be distinguished. If any classes were confounded, then power would decrease (MURANTY 1996). GÖTZ & OLLIVIER (1992) and KNOTT & HALEY (1992) using sib-pairs analysis and maximum likelihood analysis, respectively, also showed that the use of fully informative markers would greatly increase the power of QTL detection. In general, the power of backcross-type families was higher than intercross-type families. Similar results were obtained by LUO (1993) and SOLLER & GENIZI (1978).

Effect of sample size

The full-sib families were assumed to be independent, which can be thought of as a single-pair mating design structure. Increasing the number of offspring per family was found to be more efficient than increasing the number of families for a fixed total

population size (Table 3). Table 3 shows that for a given experimental size of 1000, the 5 families with 200 offspring each (5 × 200) gave higher power compared to 10 × 100 and 20 × 50 combinations. Similar results were obtained by several researchers (HILL 1975, SOLLER & GENIZI 1978, WELLER *et al.* 1990, LUO 1993, VAN DER BEEK *et al.* 1995).

MURANTY (1996) found that the power increases when variance explained by QTL and/or population size increases, and when these factors determine a low power level, the power decreases as the number of parents increases. However, at a high power level, the power increases as the number of parents increases. As a result, MURANTY (1996) suggested that the use of only one full-sib family for QTL detection is often less powerful, especially when QTL effects to be detected explain more than 10 per cent of phenotypic variance. The reason for this is that the total variance in a population attributable to QTL is better sampled with more than two parents than with only two parents. However, single full-sib families are being used for QTL mapping studies, for example, in eucalyptus (GRATTAPAGLIA *et al.* 1995) and loblolly pine (KNOTT *et al.* 1997).

Effect of gene action and allele frequency

In our study we evaluated the effect of additive and non-additive QTL effects on the power of linkage detection. REBAÏ & GOFFINET (1993) suggested that at the QTL detection step, it is better to neglect dominance if it is not very large. However, a recent study by LI *et al.* (1996) reported that the dominance variance contributes significantly to variation in tree height and diameter in loblolly pine. At $p = 0.5$, the power of PIC and FI family type designs increases as the dominance ratio increase. However, at $p = 0.5$, there was almost no effect of dominance on the power of PBC-type family design (Table 5). Similar results were reported by LUO (1993) for PIC and PBC-type family designs. Table 5 also showed that power of QTL detection using FI markers is greater compared to other two marker designs, at different levels of dominance ratio. For an additive gene action, we also evaluated the power of linkage detection at different allele frequencies (Table 6). It shows that the power is highest when $p = 0.50$. The effects of gene frequency and dominance become important when the number of families is small. This is because the probability that the marker contrast in each family be zero is so

large that even an infinite number of offspring will not meet the power requirement (SOLLER & GENIZI 1978). This effect is generally unimportant except when a dominant allele is also the more frequent. However, with the method presented in our paper, it is likely that those families with zero contrast will nevertheless contribute to the significance of the variance between marker types within families. Thus, the loss in power due to probability of sampling families with zero marker contrast can be reduced.

In this study, only bi-allelic QTL was considered. The use of fully-informative markers permits the assessment of multiple-allele QTL. Evidence of existence of more than two QTL alleles has been reported in loblolly pine (GROOVER *et al.* 1994). Degree of dominance must be estimated separately from the original QTL analysis. Pedigree and population level studies are needed to determine the prevalence of multiple-allele QTL (WILLIAMS 1996). However, the levels of power obtained for detection of linkage between a FI marker and a QTL, bi-allelic or multi-allelic, are quite similar under the given conditions (MURANTY 1996).

Comparison of the power of three marker designs revealed that FI marker design was more powerful than PBC and PIC deigns. It would be quite useful to consider using information from the whole population rather than subsets of it (i.e., combining PIC, PBC and FI family types). The joint analysis of any informative family types can be done following the suggestions of KNOTT (1994). Many alleles in a population are necessary to obtain a fully-informative marker for crosses among several parents. Isoenzymes and restricted fragment length polymorphisms (RFLP) have seldom met this criteria, but micro-satellite (SSR) and expressed sequence tag (EST) techniques promise to provide enough alleles and are currently being developed in *Pinus radiata.*

ACKNOWLEDGEMENTS

S Kumar thanks N. Z. Forest Research Institute, Rotorua, New Zealand, for financial support. We thank Drs J. I. WELLER, R. D. BURDON and Z. W. LUO for helpful discussion.

REFERENCES

CARBONELL, E. A., GERIG, T. M., BALANSARD, E. & ASINS, M. J. 1992: Interval mapping in the analysis of nonadditive quantitative trait loci. *Biometrics* **48**:305-315.

FALCONER, D. S. 1989: Introduction to Quantitative Genetics. Longman, Scientific & Technical, London, 438 pp.

GÖTZ, K. W. & OLLIVIER, L. 1992: Theoretical aspects of applying sib-pair linkage tests to livestock species. *Genetics Selection Evolution* **24**:29-42.

GRATTAPAGLIA, D., BERTOLUCCI, F. L. & SEDEROFF, R. R. 1995: Genetic mapping of QTL controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. *Theoretical and Applied Genetics* **90**:933-947.

GROOVER, A., DEVEY, M., FIDDLER, T., LEE, J., MEGRAW, R., MITCHEL-OLDS, T., SHERMAN, B., VUJCIC, S., WILLIAMS, C. & NEALE, D. 1994: Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**:1293-1300.

HILL, A. P. 1975: Quantitative linkage: a statistical procedure for its detection and estimation. *Annals of Human Genetics* **38**:439-449.

JAYAKAR, S. D. 1970: On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics* **26**:451-464.

KNOTT, S. A. 1994: Prediction of the power of detection of marker-quantitative trait locus linkages using analysis of variance. *Theoretical and Applied Genetics* **89**:318-322.

KNOTT, S. A. & HALEY, C. S. 1992: Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* **132**:1211-1222.

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.

LI, B., MCKEAND, S. E. & WEIR, R. J. 1996: Genetic parameter estimates and selection efficiency for the loblolly pine breeding in south-eastern U. S. *In*: Proceedings QFRI-IUFRO conference on Tree Improvement for Sustainable Tropical

Forestry, Caloundra, Queensland, Australia, 27 October-1 November 1996, p. 164-168.

LUO, Z. W. 1989: Polygene location and selection for heterotic traits in *Drosophila melanogaster*. Ph D thesis, Univ. Birmingham, UK.

LUO, Z. W. 1993: The power of two experimental designs for detecting linkage between a marker locus and a locus affecting a quantitative character in a segregating population. *Genetics Selection Evolution* **25**:249-261.

LYNCH, M. & WALSH, B. 1997: Genetics and Analysis of Quantitative Traits. Sinauer Associates, Massachusetts, USA. 980 p.

MURANTY, H. 1996: Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**:156-165.

REBAÏ, A. & GOFFINET, B. 1993: Power of tests for QTL detection using replicated progenies derived from a diallel cross. *Theoretical and Applied Genetics* **86**:1014-1022.

SAS INSTITUTE INC. 1986: "SAS User's Guide: Statistics", Version 5 edition. SAS Institute Inc., Cary, NC. 956 pp.

SOLLER, M. & GENIZI, A. 1978: The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* **34**:47-55.

VAN DER BEEK, S., VAN ARENDONK, J. A. M. & GROEN, A. F. 1995: Power of two-and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theoretical and Applied Genetics* **91**:1115-1124.

WELLER, J. I., KASHI, Y. & SOLLER, M. 1990: Power of daughter and granddaughter designs for determining linkage between marker locus and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**:2525-2537.

WILLIAMS, C. G. 1996: How will genomic mapping shape forest tree breeding strategy? *In*: Proceedings QFRI-IUFRO conference on Tree Improvement for Sustainable Tropical Forestry, Caloundra, Queensland, Australia, 27 October-1 November 1996, p. 464-466.

Appendix 1. Expected values (E_{ij}) of the quantitative trait value (y) within observed marker genotype (M_iM_j) of offspring within sibship. Parental marker genotypes are: M_1M_2 x M_3M_4. The cross represents all possible parental genotypes at the QTL along with their probabilities (Prob.). It is assumed that $Q_1Q_1 \sim N(a, \sigma^2)$, $Q_1Q_2 \sim N(d, \sigma^2)$, $Q_2Q_2 \sim N(-a, \sigma^2)$. The p and q represents QTL alleles frequencies and r ($s = 1 - r$) is the recombination rate between marker and the QTL.

Cross	Prob.	E_{13}	E_{14}	E_{23}	E_{24}
$Q_1Q_1 \times Q_1Q_1$	p^4	a	a	a	a
$Q_1Q_1 \times Q_1Q_2$	p^3q	$a - r(a-d)$	$d + r(a-d)$	$a - r(a-d)$	$d + r(a-d)$
$Q_1Q_1 \times Q_2Q_1$	p^3q	$d + r(a-d)$	$a - r(a-d)$	$d + r(a-d)$	$a - r(a-d)$
$Q_1Q_1 \times Q_2Q_2$	p^2q^2	d	d	d	d
$Q_1Q_2 \times Q_1Q_1$	p^3q	$a - r(a-d)$	$a - r(a-d)$	$d + r(a-d)$	$d + r(a-d)$
$Q_1Q_2 \times Q_1Q_2$	p^2q^2	$(s^2 - r^2)a + 2rsd$	$(s^2 + r^2)d$	$(s^2 + r^2)d$	$(r^2 - s^2)a + 2rsd$
$Q_1Q_2 \times Q_2Q_1$	p^2q^2	$(s^2 + r^2)d$	$(s^2 - r^2)a + 2rsd$	$(r^2 - s^2)a + 2rsd$	$(s^2 + r^2)d$
$Q_1Q_2 \times Q_2Q_2$	pq^3	$d - r(a + d)$	$d - r(a + d)$	$-a + r(d + a)$	$-a + r(d + a)$
$Q_2Q_1 \times Q_1Q_1$	p^3q	$d + r(a - d)$	$d + r(a - d)$	$a - r(a-d)$	$a - r(a-d)$
$Q_2Q_1 \times Q_1Q_2$	p^2q^2	$(s^2 + r^2)d$	$(r^2 - s^2)a + 2rsd$	$(s^2 - r^2)a + 2rsd$	$(s^2 + r^2)d$
$Q_2Q_1 \times Q_2Q_1$	p^2q^2	$(r^2 - s^2)a + 2rsd$	$(s^2 + r^2)d$	$(s^2 + r^2)d$	$(s^2 - r^2)a + 2rsd$
$Q_2Q_1 \times Q_2Q_2$	pq^3	$-a + r(d + a)$	$-a + r(d + a)$	$d - r(a + d)$	$d - r(a + d)$
$Q_2Q_2 \times Q_1Q_1$	p^2q^2	d	d	d	d
$Q_2Q_2 \times Q_1Q_2$	pq^3	$d - r(a + d)$	$-a + r(a + d)$	$d - r(a + d)$	$-a + r(a + d)$
$Q_2Q_2 \times Q_2Q_1$	pq^3	$-a + r(a + d)$	$d - r(a + d)$	$-a + r(a + d)$	$d - r(a + d)$
$Q_2Q_2 \times Q_2Q_2$	q^4	$-a$	$-a$	$-a$	$-a$

Appendix 2. Variance [*Var (i,j)*] of the quantitative trait value (y) within observed marker genotype (M_iM_j) of offspring within sibship. Parental marker genotypes are: M_1M_2 x M_3M_4. The cross represents all possible parental genotypes at the QTL along with their probabilities (Prob.). It is assumed that $Q_1Q_1 \sim N(a, \sigma^2)$, $Q_1Q_2 \sim N(d, \sigma^2)$, $Q_2Q_2 \sim N(-a, \sigma^2)$. The p and q represents QTL alleles frequencies and r ($s = 1 - r$) is the recombination rate between marker and the QTL.

Cross	Prob.	*Var (1,3)*	*Var (1,4)*	*Var (2,3)*	*Var (2,4)*
$Q_1Q_1 \times Q_1Q_1$	p^4	σ^2	σ^2	σ^2	σ^2
$Q_1Q_1 \times Q_1Q_2$	p^3q	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$
$Q_1Q_1 \times Q_2Q_1$	p^3q	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$
$Q_1Q_1 \times Q_2Q_2$	p^2q^2	σ^2	σ^2	σ^2	σ^2
$Q_1Q_2 \times Q_1Q_1$	p^3q	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$
$Q_1Q_2 \times Q_1Q_2$	p^2q^2	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(1-2r)]$	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(2r-1)]$
$Q_1Q_2 \times Q_2Q_1$	p^2q^2	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(1-2r)]$	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(2r-1)]$	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$
$Q_1Q_2 \times Q_2Q_2$	pq^3	$\sigma^2 + rs\,(d + a)^2$	$\sigma^2 + rs\,(d + a)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$
$Q_2Q_1 \times Q_1Q_1$	p^3q	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$	$\sigma^2 + rs\,(a - d)^2$
$Q_2Q_1 \times Q_1Q_2$	p^2q^2	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(2r-1)]$	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(1-2r)]$	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$
$Q_2Q_1 \times Q_2Q_1$	p^2q^2	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(2r-1)]$	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$	$\sigma^2 + 2\,rsa^2 + d^2(r^2 + s^2)[1 - (r^2 + s^2)]$	$\sigma^2 + 2rs\,[a^2 + d^2(1-2rs)-2da(1-2r)]$
$Q_2Q_1 \times Q_2Q_2$	pq^3	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$
$Q_2Q_2 \times Q_1Q_1$	p^2q^2	σ^2	σ^2	σ^2	σ^2
$Q_2Q_2 \times Q_1Q_2$	pq^3	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$
$Q_2Q_2 \times Q_2Q_1$	pq^3	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$	$\sigma^2 + rs\,(a + d)^2$
$Q_2Q_2 \times Q_2Q_2$	q^4	σ^2	σ^2	σ^2	σ^2

Appendix 3. Mean and variances between marker genotype classes within sibships. Parental marker genotypes are: M_1M_2 x M_3M_4. The cross represents all possible parental genotypes at the QTL along with their probabilities (Prob.), E_{ij} represents the expected value of offspring having marker genotype M_iM_j. It is assumed that $Q_1Q_1 \sim N(a, \sigma^2)$, $Q_1Q_2 \sim N(d, \sigma^2)$, $Q_2Q_2 \sim N(-a, \sigma^2)$. The p and q represent QTL alleles frequencies and r ($s = 1 - r$) is the recombination rate between marker and the QTL.

Cross	Prob.	$\frac{1}{4}(E_{13} + E_{14} + E_{23} + E_{24})$	Variance between E_{13}, E_{14}, E_{23}, E_{24}
$Q_1Q_1 \times Q_1Q_1$	p^4	a	0
$Q_1Q_1 \times Q_1Q_2$	p^3q	$\frac{1}{2}(a + d)$	$\frac{1}{4}(s - r)^2(a - d)^2$
$Q_1Q_1 \times Q_2Q_1$	p^3q	$\frac{1}{2}(a + d)$	$\frac{1}{4}(s - r)^2(a - d)^2$
$Q_1Q_1 \times Q_2Q_2$	p^2q^2	d	0
$Q_1Q_2 \times Q_1Q_1$	p^3q	$\frac{1}{2}(a + d)$	$\frac{1}{4}(s - r)^2(a - d)^2$
$Q_1Q_2 \times Q_1Q_2$	p^2q^2	$\frac{1}{2}d$	$\frac{1}{2}[(s - r)^2a^2 + (r^2+s^2)^2d^2 + 4r^2s^2d^2] - \frac{1}{4}d^2$
$Q_1Q_2 \times Q_2Q_1$	p^2q^2	$\frac{1}{2}d$	$\frac{1}{2}[(s - r)^2a^2 + (r^2+s^2)^2d^2 + 4r^2s^2d^2] - \frac{1}{4}d^2$
$Q_1Q_2 \times Q_2Q_2$	pq^3	$\frac{1}{2}(d - a)$	$\frac{1}{4}(a + d)^2(s - r)^2$
$Q_2Q_1 \times Q_1Q_1$	p^3q	$\frac{1}{2}(a + d)$	$\frac{1}{4}(s - r)^2(a - d)^2$
$Q_2Q_1 \times Q_1Q_2$	p^2q^2	$\frac{1}{2}d$	$\frac{1}{2}[(s - r)^2a^2 + (r^2+s^2)^2d^2 + 4r^2s^2d^2] - \frac{1}{4}d^2$
$Q_2Q_1 \times Q_2Q_1$	p^2q^2	$\frac{1}{2}d$	$\frac{1}{2}[(s - r)^2a^2 + (r^2+s^2)^2d^2 + 4r^2s^2d^2] - \frac{1}{4}d^2$
$Q_2Q_1 \times Q_2Q_2$	pq^3	$\frac{1}{2}(d - a)$	$\frac{1}{4}(a + d)^2(s - r)^2$
$Q_2Q_2 \times Q_1Q_1$	p^2q^2	d	0
$Q_2Q_2 \times Q_1Q_2$	pq^3	$\frac{1}{2}(d - a)$	$\frac{1}{4}(a + d)^2(s - r)^2$
$Q_2Q_2 \times Q_2Q_1$	pq^3	$\frac{1}{2}(d - a)$	$\frac{1}{4}(a + d)^2(s - r)^2$
$Q_2Q_2 \times Q_2Q_2$	q^4	$-a$	0

Chapter 3

MARKER-QTL LINKAGE DETECTION IN SELF-FAMILIES OF OUTBRED POPULATIONS

S. Kumar, R. D. Burdon & D. J. Garrick

SILVAE GENETICA (1999) 48: 227-234

ABSTRACT

In forest trees which are normally outcrossing, inbreeding by self-fertilisation (selfing) generally has deleterious effects including reduced seed set, poor seed germination, and slow seedling growth. Inbreeding depression (ID) is mainly caused by deleterious alleles that will be almost never expressed under panmixis. Until the advent of molecular markers, there has been no way to track most of the individual genes causing ID. In this study, the theory for a single-marker ANOVA method was developed to find the linkage between a marker locus and a gene causing ID in growth traits in self-families of outbred populations. The power of linkage detection, which was at the lower limit because of single-marker method, was calculated for a wide range of progeny sizes and genetic parameters at the quantitative trait locus (QTL). The magnitude of the gene effect was found to have an enormous effect on the power. The situations where the QTL detected in a self-family can be considered as those expressed in normal course of outbreeding are also discussed.

Key words: Selfing, inbreeding, molecular marker, QTL, outbred.

INTRODUCTION

Inbreeding, which is reduction in heterozygosity across the genome resulting from mating among relatives including selfing, usually affects the phenotypic performance of inbred offspring. The deleterious effect of inbreeding on the phenotype is termed as inbreeding depression. In outbreeding forest trees, inbreeding by self-fertilisation (selfing) generally has highly deleterious effects which include reduced seed set, poor seed germination, slow seedling growth and abnormal morphology (WILLIAMS & SAVOLAINEN, 1996). Inbreeding depression (ID) is a complex quantitative phenomena, presumably controlled by many deleterious genes of different magnitudes of effects. Inbreeding depression is common and severe in many tree species, particularly conifers which are believed to have large numbers of recessive embryo lethals and post-germination lethals.

There are various genes affecting components of fitness such as viability and vigour/growth. Most reports of ID in conifers centre on embryo-stage lethals and other

deleterious recessives which affect seedlings. The number of reduced filled seed upon selfing is attributed to embryo mortality because they are homozygous for lethals or deleterious genes which act at early stages of development. The reduction in viability upon inbreeding is well-studied, and number of lethals equivalents have been estimated for various species (WILLIAMS & SAVOLAINEN, 1996). Many studies have reported that selfed seedlings have higher mortality than outcrossed seedlings (YADZANI *et al.*, 1985; PLESSAS & STRAUSS, 1986; MUONA *et al.*, 1987). Selfed progeny also exhibit strong ID for growth at early ages (FRANKLIN, 1970 and 1972; WILCOX, 1983).

From the seedling stage onward, many quantitative characters such as growth are considered. The effect of ID on growth traits has been studied widely. In this study, we will use the term "performance genes" to refer to the genes which are expressed in fully viable (but not necessarily very fit) genotypes. Performance genes should play an important role in expression of ID since strong ID usually occurs at least in later stage of the life cycle of many plants (see CHARLESWORTH & CHARLESWORTH, 1987; HUSBAND & SCHEMSKE, 1996). WILCOX (1983) reported no ID for wood density and bole straightness in selfed offspring of some parents (measured at 7-8 years) of radiata pine (*Pinus radiata* D. Don). However, other growth traits like height and diameter showed severe ID. One study on Norway spruce (*Picea abies* L.) showed little ID (6%) for wood density at age 10 years (SKRØPPA, 1996).

Several authors have discussed the possibilities of using selfing as a breeding tool in forest trees (BARKER & LIBBY, 1974; LINDGREN, 1975; WILCOX, 1983; WILLIAMS & SAVOLAINEN, 1996). The advent of specialty populations bred specifically for quality and for disease resistance has created a great interest in studies on ID in these traits. Multi-generation studies of inbreeding are required to study the genetic basis of ID and especially to test whether it is possible to eliminate or purge deleterious alleles (WILLIAMS & SAVOLAINEN, 1996). Purging of deleterious alleles will be more effective and efficient if the alleles can be associated with molecular markers so that marker-aided-selection (MAS) may be used to identify and eliminate the deleterious alleles (KUANG *et al.*, 1998). However, one must use large numbers of offspring per self-fertile parent to select strongly and effectively against loss of vigour or other performance

indicators. Juvenile traits and markers linked to older-tree traits could be selected in the first-stage culling (WILLIAMS & NEALE, 1992).

DETECTING MARKER-QTL LINKAGE

Using molecular markers it is now relatively straightforward to create a genetic map of the genome to identify loci affecting quantitative traits (QTLs) of interest. Detection of QTLs in self-families, however, is not the same as in outbred pedigrees. There can be up to four alleles segregating at a locus in an outbred full-sib pedigree. It would result in various mating type configurations (e.g. backcross, intercross and non-informative types) with respect to a marker locus. Power of QTL detection experiments in full-sib pedigrees of outbred forest trees has been studied theoretically (e.g., MURANTY, 1996; KUMAR *et al.*, 2000). An informative selfed pedigree has a two-allele 'intercross' mating type configuration at a locus. The other major difference is regarding the specific QTLs being detected. Some of the QTLs in self-families will represent genetic load; however, there will be some individuals without appreciable load so the QTLs detected in these self-pedigrees would be similar to those expressed under the normal course of outbreeding (PLOMION *et al.*, 1996).

Molecular markers are increasingly used in self-families of outbred organisms to find genes causing inbreeding depression in viability and growth traits (HEDRICK & MUONA, 1990; FU & RITLAND, 1994a,b; PLOMION *et al.*, 1996; KUANG *et al.*, 1998). While mapping performance loci involves the comparisons of quantitative trait means among marker genotypes, mapping viability loci involves analysis of marker genotype frequencies. Different analytical methods are required to characterise genes affecting various fitness-related components such as viability and growth.

HEDRICK & MUONA (1990) used a single-marker approach to detect and characterise viability alleles in Scots pine (*Pinus sylvestres* L.). FU & RITLAND (1994a) studied the statistical properties of mapping recessive viability loci using the single-marker approach. Also, they showed that two flanking markers provide vastly superior estimation properties and reduced sample sizes compared to those required by a single marker. KUANG *et al.* (1998) used a single marker approach to find an allele responsible for seedling death in radiata pine. To detect viability genes, segregating families are

analysed by chi-square tests for deviations from the Mendelian segregation ratio 1:2:1 at a codominant marker locus.

FU & RITLAND (1994b) used a single-marker analysis of variance (ANOVA) approach to draw inferences about performance genes (labelled "fecundity" genes by the authors) contributing to ID in fitness (growth) traits from an experimental data containing two self-families of *Mimulus guttatus* DC (Scrophulariaceae). PLOMION *et al.* (1996), using experimental data from a self-family of a hybrid tree of maritime pine, showed a higher efficiency of interval mapping compared to single-marker ANOVA approach for detecting QTLs for a growth trait (height). The expected power of interval mapping (LANDER & BOTSTEIN, 1989) or multiple-marker mapping (KNOTT *et al.*, 1996) will certainly exceed that from single-marker approach. Thus, the power of linkage detection calculated from single-marker approach will provide the lower limit. Except for a few experimental studies, not much theory has been developed to quick-screen different types of experimental designs (or specifically, size of mapping populations) required for finding linkage between a marker locus and a performance-trait locus in self-families of outbred populations.

The objective of this study was to develop theory for a single-marker ANOVA method for finding the linkage between a codominant marker locus and a performance locus in self-families of outbred populations. This paper re-examines the utility of selfing, in the light of available molecular marker technology, in breeding programmes of outbred species. This paper also discusses the situations where the QTLs detected in self-families can be interpreted as QTLs that are expressed in the normal course of outbreeding.

Basic Assumptions and Statistical Model

The method involves analysing progeny from self-families in a normally outbreeding population. Two autosomal loci are considered; one of them affects a quantitative trait (QTL) while the other is a codominant marker. The marker and the QTL are assumed to be in linkage equilibrium. The two loci are linked with a recombination frequency of r (s = 1 - r). Let the frequency of allele Q_1 at the QTL be denoted as p (q = 1 - p), and the phenotypic distributions of different genotypes at the QTL i.e., Q_1Q_1, Q_1Q_2 and Q_2Q_2

are assumed to be normally distributed with means: a, d, -a, respectively and common variance σ^2. The a and d represent the additive and dominance effects at the QTL. With just one QTL and no other gene effects, σ^2 will be the environmental variance, whereas in the presence of unlinked QTLs, it will also include 'background' genetic variance at these loci (e.g., polygene variance).

Parents are assumed to be heterozygous at the marker locus. Since the parent is being selfed, the marker-QTL genotypes of male and female parents are identical. Let the parental genotype at the marker locus be M_1M_2 and three marker genotype classes (M = 3) are distinguishable in the selfed progeny: M_1M_1, M_1M_2, and M_2M_2 segregating with a 1:2:1 ratio. Let n_{ij} denote the number of sibs within the j^{th} marker class within the i^{th} self-family. Also each family (N_f) has a constant size of N_o and thus the total experimental size is $N_f \times N_o$.

In carrying out a simultaneous analysis of several self-families, it is necessary to take into account that linkage relationship between a marker and a performance locus will differ among different individuals. A hierarchical ANOVA can be applied which allows marker effects to change sign over sibships. The linear model for the phenotype of the quantitative trait measured on the k^{th} sib (k = 1, 2, . . ., n_{ij}) with the j^{th} marker genotype (j = 1, 2, . . ., M) within the i^{th} family (i = 1, 2, . . ., N_f) can be written as:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk} \qquad (1)$$

where μ is an overall mean, α_i, β_{ij} and e_{ijk} are contributions from the family, from the marker genotype within family and a random contribution of environment to the individual, respectively. They are assumed to be independently and normally distributed with zero means and variances σ_α^2, σ_β^2 and σ_e^2, respectively. The ANOVA for this model is given in Table 1.

Under the assumption of a constant size of sibship (N_o) and

$$n_{i1} : n_{i2} : n_{i3} \approx 1 : 2 : 1, \qquad (2)$$

the approximation for n_0 will be : $n_0 \approx (5/16)\, N_o$. (3)

Table 1. ANOVA for a two-factor completely nested design.

Source	Degrees of freedom	MS	EMS
Between families	$N_f - 1$	MS_s	-
Between marker genotypes within families	$\sum(M_i - 1)$	MS_m	$\sigma_e^2 + n_0\sigma_\beta^2$
Within marker genotype within families	$\sum(n_{ij} - 1)$	MS_w	σ_e^2

All possible marker-QTL genotypes of parents and the gametes inherited by the offspring are given in Table 2 with their probabilities. Using these probabilities, the expected values of the quantitative trait value, y, were obtained for different marker genotypes of offspring (Table 3).

Table 2. Probabilities of various gametes inherited from parents to progeny. Recombination rate between marker and QTL is r.

Parental genotype	Gametes			
	M_1Q_1	M_1Q_2	M_2Q_1	M_2Q_2
M_1Q_1 / M_2Q_1	1/2	0	1/2	0
M_1Q_1 / M_2Q_2	(1 - r) / 2	r/2	r/2	(1 - r)/2
M_1Q_2 / M_2Q_1	r/2	(1 - r)/2	(1 - r)/2	r/2
M_1Q_2 / M_2Q_2	0	1/2	0	1/2

Table 3. Expected values of quantitative trait value (y) within families within marker genotypes. Assuming that $Q_1Q_1 \sim N(a, \sigma^2)$, $Q_1Q_2 \sim N(d, \sigma^2)$, $Q_2Q_2 \sim N(-a, \sigma^2)$. The cross represents all possible parental genotypes at the QTL along with their probabilities (Prob), E_{ij} represents the expected value of offspring having marker genotype M_iM_j. The p and q represent QTL alleles frequencies and r (s = 1 - r) is the recombination rate between marker and the QTL.

Cross	Prob	E_{11}	E_{12}	E_{22}
Q1Q1	p^2	a	a	a
Q1Q2	pq	a (1 - 2r) + 2rsd	$(r^2 + s^2)$d	-a (1 - 2r) + 2rsd
Q2Q1	pq	-a (1 - 2r) + 2rsd	$(r^2 + s^2)$d	a (1 - 2r) + 2rsd
Q2Q2	q^2	-a	-a	-a

Similarly, the variances of the trait value within families within marker genotypes were derived (Table 4). Finally, the variances between marker genotype classes within families were obtained (Table 5).

Table 4. Variance of the quantitative trait value (y) within families within marker genotypes. Assuming that $Q_1Q_1 \sim N(a, \sigma^2)$, $Q_1Q_2 \sim N(d, \sigma^2)$, $Q_2Q_2 \sim N(-a, \sigma^2)$. The cross represents all possible parental genotypes at the QTL. The probability of each cross is similar to that given in Table 3. V_{ij} represents variance within marker genotype M_iM_j. The p and q represent QTL alleles frequencies and r (s = 1 - r) is the recombination rate between marker and the QTL.

Cross	V_{11}	V_{12}	V_{22}
Q1Q1	σ^2	σ^2	σ^2
Q1Q2	$\sigma^2 + 2rs[a^2+d^2(1 - 2rs) - 2ad(1 - 2r)]$	$\sigma^2 + 2rs[a^2 + d^2(1-2rs)]$	$\sigma^2 + 2rs[a^2+d^2(1 - 2rs) + 2ad(1 - 2r)]$
Q2Q1	$\sigma^2 + 2rs[a^2+d^2(1 - 2rs) + 2ad(1 - 2r)]$	$\sigma^2 + 2rs[a^2 + d^2(1-2rs)]$	$\sigma^2 + 2rs[a^2+d^2(1 - 2rs) - 2ad(1 - 2r)]$
Q2Q2	σ^2	σ^2	σ^2

Table 5. Mean and variances between marker genotype classes within families. Assuming that $Q_1Q_1 \sim N(a, \sigma^2)$, $Q_1Q_2 \sim N(d, \sigma^2)$, $Q_2Q_2 \sim N(-a, \sigma^2)$. The cross represents all possible parental genotypes at the QTL. The probability of each cross is similar to that given in Table3. E_{ij} represents the expected value of offspring having marker genotype M_iM_j. The p and q represent QTL alleles frequencies and r (s = 1 - r) is the recombination rate between marker and the QTL.

Cross	Mean = $0.25(E_{11}+2E_{12}+E_{22})$	Variance between E_{11}, E_{12}, E_{22}
Q1Q1	a	0
Q1Q2	½ d	$½(1-2r)^2a^2 + ¼(1-2r)^4d^2$
Q2Q1	½ d	$½(1-2r)^2a^2 + ¼(1-2r)^4d^2$
Q2Q2	-a	0

Using the results from Table 3, 4 and 5, the expected variances: between families (σ_α^2), between marker genotypes within-families (σ_β^2), within-marker genotype within-families (σ_e^2), and total phenotypic variance (σ_T^2) were derived as:

$$\sigma_\alpha^2 = 2pq[a^2 + 0.25d^2(1-2pq) - ad(p-q)] \quad (4)$$

$$\sigma_\beta^2 = 2pq[0.5(1-2r)^2 a^2 + 0.25(1-2r)^4 d^2] \quad (5)$$

$$\sigma_e^2 = \sigma^2 + 4pqrs[a^2 + d^2(1-2rs)] \quad (6)$$

$$\sigma_T^2 = \sigma^2 + 2pq\{a^2(1+F) + d^2(1-F)[1-(1-F)2pq] - (p-q)2ad(1-F)\} \quad (7)$$

Total genetic variance, V_G, arising from one locus (QTL) for given inbreeding coefficient (F) can be written as (KEMPTHORNE, 1973):

$$V_G = 2pq\{a^2(1+F) + d^2(1-F)[1-(1-F)2pq] - (p-q)2ad(1-F)\} \quad (8)$$

The value of the inbreeding coefficient (F) after one generation of selfing will be 0.50 in equations (7) and (8). When F = 0, the expression in equation (8) becomes the genetic variance at a locus in random mating populations. From equation (5) it can be easily shown that the expected variance between marker genotypes within families (σ_β^2) will be zero if there is no linkage between the marker and the QTL, i.e., r = s = 0.5. Under the null hypothesis (H_0: r = 0.5) the ratio MS_m / MS_w (Table 1) is distributed as a central F-variable; whereas this ratio will be a noncentral F-variable when r is less than 0.5 (JAYAKAR, 1970; LUO, 1993). Using the standard definition, the power function for linkage can be written in the following general form:

$$\text{Power} = \Pr\left[F_{(v1,\, v2;\, \delta)} > F_{(\alpha;\, v1,\, v2)}\right], \quad (9)$$

where $F_{(v1,\, v2;\, \delta)}$ is a noncentral F-variable with degrees of freedom v_1 and v_2 and noncentrality parameter δ, while $F_{(\alpha;\, v1,\, v2)}$ is the upper α point of a central F-variable with degrees of freedom v_1 and v_2. The value of noncentrality parameter, δ, was calculated as (LUO, 1993):

$$\delta = (M \times N_f - 1)\, n_0\, \sigma_\beta^2 / \sigma_e^2. \qquad (10)$$

Using different combinations of design parameters (N_f and N_o), genetic parameters at the QTL (p, dominance ratio ($f = d/a$)) and recombination frequency (r), the noncentrality parameter can be calculated. After that power can be easily calculated using (9).

The variance components derived in equations (4) to (7) are for a multiple self-families situation. For the single self-family case, the expected variance ratios for each cross type were derived following JAYAKAR (1970) and given in Table 6. These variance ratios have an expected value of 1 in the absence of linkage, and are distributed as F-variables. Table 6 shows that the excess of expected value of F-variable over 1 is given by $K(N_0 - 1)$ where K is ratio of between-marker genotype variance to within-marker genotype variance. Under the alternative hypothesis of linked QTL, these variance ratios will follow noncentral F-distribution with noncentrality parameter $K(N_0 - 1)$. Thus, for a single self-family the power of marker-QTL linkage detection was calculated separately for each cross type and then were pooled together using the probability of each cross type.

Table 6. Expected variance ratios in the single self-family with the probabilities of their occurrences.

Cross	Probability	Expected variance ratio
Q1Q1	p^2	1
Q1Q2	pq	$1+\frac{(N_0-1)(0.5(1-2r)^2a^2+0.25(1-2r)^4d^2)}{\sigma^2+2rs(a^2+d^2(1-2rs))}$
Q2Q1	pq	$1+\frac{(N_0-1)(0.5(1-2r)^2a^2+0.25(1-2r)^4d^2)}{\sigma^2+2rs(a^2+d^2(1-2rs))}$
Q2Q2	q^2	1

RESULTS AND DISCUSSION

The variance component expressions were used to calculate the power of marker-QTL linkage detection in independent self-families. Different combinations of design

parameters (N_f, N_o) along with parameters at the QTL such as dominance ratio, allele frequency, size of gene effect were used in deterministic simulation. The variance contributed, as a function of allele frequency and gene action, by a selfing locus to the breeding population is shown in Figure 1. It shows that when the gene action at the QTL is purely additive then the variance explained will be maximum only when p = q = 0.50. However, for dominant gene action, the maximum variance contributed by QTL will be at a lower allele frequency (p) of about 0.35. It can be seen that the variance contributed by a locus is higher when gene action is dominant and the frequency (p) is less than 0.70 (Figure 1). It suggests that the power of detecting linkage between a marker locus and a dominant gene would be higher compared to that of an additive gene unless the dominant gene is close to fixation (p is close to 1.0, which will tend to occur in most cases of genetic load alleles of large effects).

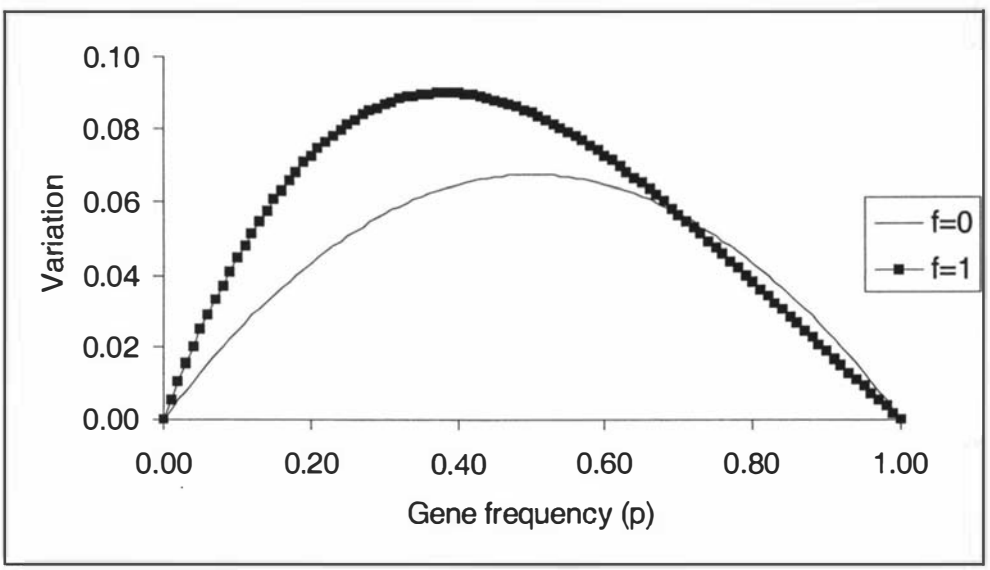
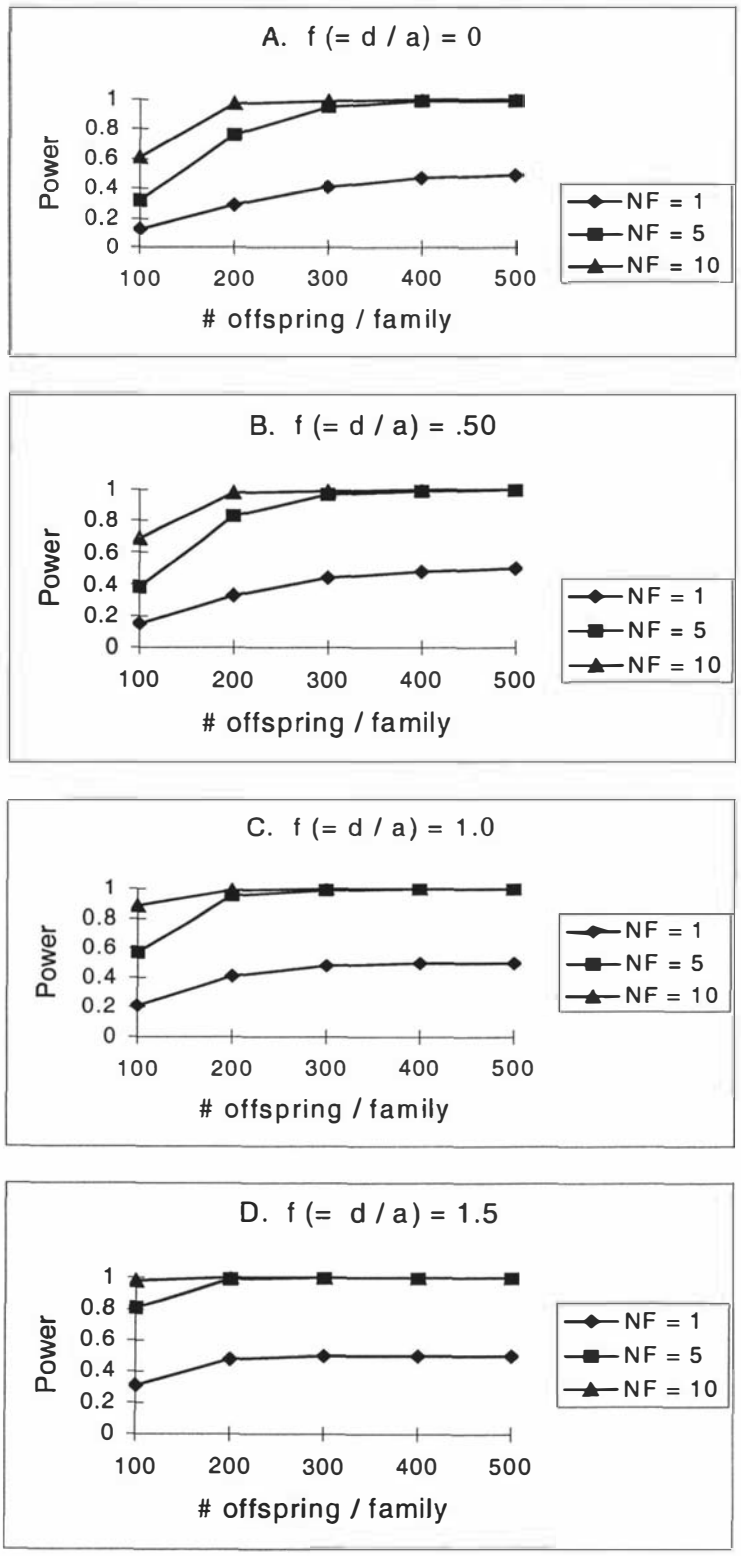


Figure 1. Change in the contribution made by a gene to the variation of a breeding population according to the frequency and gene action. Half the difference between QTL homozygotes (i.e., a) = 0.30 SD. The dominance ratio (f) = 0 and 1 represent additive and dominant gene action, respectively.

Effect of Sample Size

The effect of different levels of dominance at the QTL, on the power of linkage detection is shown in Figure 2. It shows that as the dominance increases the power also increases. However, for a single self-family the maximum achievable power is only 0.5.

Figure 2. - The power of marker-QTL linkage detection. Different gene actions considered are: **A**: Additive, **B**: Partial-dominance, **C**: Dominant, **D**: Over-dominance. The other assumptions are: p (allele frequency of dominant allele, Q_1) = 0.5, recombination rate (r) = 0.0, Type1 error rate = 0.01, half the difference between QTL homozygotes (i.e., a) = 0.30 SD and NF = number of self-families.



It also shows that for a large number of families the power of finding marker-trait association is higher compared to a single self-family. This is simply because of the high probability of finding segregating (at the QTL) families with a large number of families. Nevertheless increasing the number of offspring per family was found to be more efficient than increasing the number of families for a fixed total population size (Figure 2). Similar results have been reported by several authors (HILL, 1975; SOLLER & GENIZI, 1978; WELLER *et al.*, 1990). It might, however, be in practice quite difficult to meet the sample size requirement in self-families because of empty seeds or high seedling mortality.

Effect of Gene Action and Gene Frequency

The effect of gene action and gene frequency, on the power of linkage detection, in a multi-family situation is shown in Table 7. It shows that for a fixed size of gene effect and recombination rate the power of linkage detection is slightly increasing with the dominance ratio (f). As the QTL allele frequency departs from the intermediate frequency (0.5), the power decreases.

Table 7. The effect of gene frequency (p) and gene action (f = d/a) on the power of linkage detection in a multiple self-families situation. The half the difference between two QTL homozygotes (a) = 0.30 SD, recombination rate (r) = 0.10, and type-1 error rate = 0.01.

		f = 0.0			f = 0.50			f = 1.0			f = 1.5		
		p =			p =			p =			p =		
N_f	N_o	.25	.50	.75	.25	.50	.75	.25	.50	.75	.25	.50	.75
5	100	.10	.16	.10	.11	.18	.11	.16	.25	.15	.25	.38	.23
	300	.53	.73	.53	.59	.78	.57	.73	.89	.71	.89	.97	.87
	500	.86	.96	.86	.90	.98	.89	.96	.99	.96	.99	1.0	.99
10	100	.19	.31	.19	.22	.36	.21	.32	.49	.29	.49	.69	.45
	300	.85	.96	.85	.89	.98	.88	.96	.99	.96	.99	1.0	.99
	500	.99	.99	.99	.99	1.0	.99	.99	1.0	.99	1.0	1.0	1.0

The effect of a wide range of parameters on the power of linkage detection in a single self-family was also investigated and the results are given in Table 8. It shows that as the difference between two QTL homozygotes increases the power of linkage detection also increases. Recombination rate between marker locus and the QTL has a enormous effect on the power. When there is no recombination (r = 0.0, marker and the QTL are on the same position on the chromosome) the power of marker-QTL linkage detection is quite high compared to when the QTL is about 10 cM (r = .10) away from the marker. A dense map will improve our ability to find close linkage between marker and the QTL. Table 8 also shows that if the frequency of the QTL allele is high or low, then the power of detecting QTL will be less compared to that for intermediate frequencies.

Table 8. The effect of gene frequency (p), gene action (f = d/a) and recombination rate (r) on the power of linkage detection in a single self-family. Half the difference between two QTL homozygotes (a) were: 0.20 and 0.40 SD, and type-1 error rate = 0.01.

		f = 0				f = 1.0			
		a = .20		a = .40		a = .20		a = .40	
p	N_o	r = 0	r = .10	r = 0	r = .10	r = 0	r = .10	r = 0	r = .10
.25	100	.04	.02	.19	.11	.06	.03	.30	.16
	300	.14	.08	.37	.33	.22	.11	.38	.37
	500	.24	.15	.38	.37	.32	.20	.38	.38
.50	100	.05	.03	.27	.15	.07	.04	.39	.22
	300	.18	.10	.49	.44	.28	.14	.50	.48
	500	.32	.19	.50	.50	.43	.27	.50	.50
.75	100	.04	.02	.19	.11	.05	.03	.28	.15
	300	.14	.08	.37	.33	.21	.11	.38	.36
	500	.24	.15	.38	.37	.32	.20	.38	.38

For the dominant gene action at the QTL, the power of linkage detection is slightly higher when the QTL allele is less frequent compared to when its frequency is higher in the population. Similar results were reported by LUO (1993), however, in random mating populations.

In this study, the power of linkage detection between a codominant marker locus and a performance locus in self-families was obtained deterministically using the single-marker approach. KNOTT *et al.* (1996), using stochastic simulations for outbred half-sib pedigrees, showed that the single-marker method provides the lower limit of the power and it can be increased using multiple-marker methods. They found that the increase in power from the use of multiple markers was greatest when markers were close together and the power was intermediate. Thus, the expected powers for different scenarios shown in this study should be considered as the lower limit. The methodology presented here provides a useful tool to enable quick screening of different scenarios (experimental and genetical) deterministically before establishing a QTL mapping trial in self-families of outbred populations. However, the flanking marker methods or multiple-marker methods should be preferred to analyse the experimental data because of their higher efficiency.

Detection of linkage between a codominant marker locus and a performance locus is based on the measurements of growth traits on survivors in self-families. A molecular marker linked with 'non lethal' ID-affecting locus (performance locus) should not show segregation distortion unless it is also linked to a lethal gene affecting viability. Our study is aimed at detection of linkage between a codominant marker locus and a performance locus and thus, the theory developed here is based on the assumption that the three marker genotypes are segregating in the ratio of 1:2:1. Selection of a locus affecting fitness trait (e.g., growth trait) may occur with or without segregation distortion. FU & RITLAND (1994b) found performance genes contributing to the ID in vigour traits in self-families of *M. guttatus*. The observed frequencies of marker genotypes at about half of the marker loci linked to the performance genes were different from the expected frequencies (1:2:1). Generally, the power to test means is greatest when sample sizes of each mean are equal (Personal communication with Professor KERMIT RITLAND). We have not studied, in our paper, about the magnitude of effect of segregation distortion on the power of detecting linkage between a marker locus and a performance locus. In the progeny of non-inbred parents, the power of QTL detection was shown to be less when there is a distorted segregation ratio (SCHAFER-PREGAL *et al.*, 1996).

IMPLICATIONS FOR BREEDING

Selfing in outcrossing forest trees usually results in inbreeding depression, with much reduced seed set and vigour. However, the use of selfs for progeny testing in normally-outcrossing forest tree species can in principle be highly efficient, and has been advocated by several researchers (BARKER & LIBBY, 1974; LINDGREN, 1975; WRIGHT, 1980; WILLIAMS & SAVOLAINEN, 1996). Selfing and sib-mating as a breeding tool has been revived because of the growing interest in small elite breeding populations. Performance of self-families can be a reliable indicator of general combining ability (GCA) under outcrossing. WILCOX (1983) noted that some families showed negligible load such that for those particular families self-family performance will give a good guide to breeding values. In the face of the parental differences in ID, self-family information could be used to give reasonable estimates of breeding value by culling the data from all but the strongest individuals within families (BARKER & LIBBY, 1974).

Performance genes contributing to ID can be found more reliably and efficiently by conducting parallel studies on self and outcross families of the same parents. The simplest option might be to use pair-crosses in conjunction with self-families. Once molecular markers linked to performance genes contributing to ID are found, the mode of gene action can be studied to help understand the genetic basis of ID (e.g., FU & RITLAND, 1994b). Using these marker-trait associations, self-families can be culled for inferior individuals, at a very early stage (depending on the age at which genetic load alleles are expressed) ahead of phenotypic expression, which are thus identified as probable homozygotes for deleterious recessive genes. However, one must use a large number of offspring per self-fertile parent to select strongly against loss of vigour (WILLIAMS & SAVOLAINEN, 1996). After culling for inferior individuals, the performance of self-families would be a reliable indicator of GCA under outcrossing. Also, the genetic correlation approach of BURDON & RUSSELL (1998) can be used as an indication of inherent reliability within a population of self-family information as a guide to parental breeding value. In the absence of non-additive genetic variance, we can expect a perfect genetic correlation between self-performance and outcrossed breeding value (GCA).

ID is the reduction in fitness and thus if some fitness trait (e.g., height, diameter etc.) shows an association with a marker in selfed progenies, then the linked QTL can be called as ID-related. However, PLOMION *et al.* (1996) used a self-family of a hybrid tree of maritime pine for mapping QTLs (ID-independent) for early seedling growth because there was no evidence of genetic load in this outcrossed tree. The absence of genetic load in their study could be considered as a fortuitous, essentially stochastic effect. However, parents that show no ill-effects from selfing and also exhibit high performance for traits of direct economic value may be used for mapping QTLs that are expressed in the normal course of outbreeding. There are other traits not obviously associated with fitness (e.g., leaf shape, flower colour etc.). The QTL found for these traits using selfed progenies would be ID-independent and can be considered as those expressed in normal course of outbreeding

ACKNOWLEDGEMENTS

We thank Dr Phillip Wilcox for comments on the earlier version of the manuscript. S. Kumar thanks ***Forest Research***, New Zealand for financial support.

REFERENCES

BARKER, J. E. & LIBBY, W. J. 1974: The use of selfing in selection of forest trees. *Journal of Genetics* **61**:152-168.

BURDON, R. D. & RUSSELL, J. H. 1998: Inbreeding depression in selfing experiments: Statistical issues. *Forest Genetics* **5**:179-189.

CHARLESWORTH, D. & CHARLESWORTH, B. 1987: Inbreeding depression and its evolutionary consequences. *Annual Review of Ecological Systems* **18**:237-268.

FRANKLIN, E. C. 1970: Survey of mutant forms. *USDA Forest Service Research Paper* SE-61. 21 p.

FRANKLIN, E. C. 1972: Genetic load in loblolly pine. *American Naturalist* **106**:262-265.

FU, Y. B. & RITLAND, K. 1994a: On estimation of linkage of marker genes to viability genes controlling inbreeding depression. *Theoretical and Applied Genetics* **88**: 925-932.

FU, Y. B. & RITLAND, K. 1994b: Marker-based inferences about fecundity genes contributing to inbreeding depression in *Mimulus guttatus*. *Genome* **37**:1005-1010.

HEDRICK, P. W. & MUONA, O. 1990: Linkage of viability genes to marker loci in selfing organisms. *Heredity* **64**:67-72.

HILL, A. P. 1975: Quantitative linkage: a statistical procedure for its detection and estimation. *Annals of Human Genetics* **38**:439-449.

HUSBAND, B. C. & SCHEMSKE, D. W. 1996: Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**:54-70.

JAYAKAR, S. D. 1970: On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics* **26**:451-464.

KEMPTHORNE, O. 1973: *An Introduction to Genetic Statistics*. Iowa State University Press, Ames IA, USA. 545 p.

KNOTT, S. A., ELSEN, J. M. & HALEY, C. S. 1996: Methods for multiple marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**:71-80.

KUANG, H., RICHARDSON, T. E., CARSON, S. D. & BONGARTEN, B. C. 1998: An allele responsible for seedling death in *Pinus radiata* D. Don. *Theoretical and Applied Genetics* **96**:640-644.

KUMAR, S., CARSON, S. D. & GARRICK, D. J. 2000: Detecting linkage between a fully-informative marker locus and a trait locus in outbred populations using analysis of variance. *Forest Genetics* **7**:47-56.

LANDER, E. S. & BOTSTEIN, D. 1989: Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**:185-199.

LINDGREN, D. 1975: Use of selfed material in forest tree improvement. Department of Forest Genetics, Royal College of Forestry, Research Notes No. 15. 76 p. Stockholm, Sweden.

LUO, Z. W. 1993: The power of two experimental designs for detecting linkage between a marker locus and a locus affecting a quantitative character in a segregating population. *Genetics Selection Evolution* **25**: 249-261.

MUONA, O., YAZDANI, R. & RUDIN, D. 1987: Genetic changes between life stages in *Pinus sylvestris*: Allozyme variation in seeds and planted seedlings. *Silvae Genetica* **36**:39-42.

MURANTY, H. 1996: Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**:156-165.

PLESSAS, M. E. & STRAUSS, S. H. 1986: Allozyme differentiation among populations, stands and cohorts in Monterey pine. *Canadian Journal of Forest Research* **16**:1155-1164.

PLOMION, C., DUREL, C.-E. & O'MALLEY, D. M. 1996: Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. *Theoretical and Applied Genetics* **93**:849-858.

SCHAFER-PREGL, R., SALAMINI, S. & GEBHARDT, C. 1996: Models for mapping quantitative trait loci (QTL) in progeny of non-inbred parents and their behaviour in presence of distorted segregation ratios. *Genetical Research* Cambridge **67**:43-54.

SKROPPA, T. 1996: Diallel crosses in *Picea abies*. II. Performance and inbreeding depression of selfed families. *Forest Genetics* **3**:69-79.

SOLLER, M. & GENIZI, A. 1978: The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* **34**:47-55.

WELLER, J. I., KASHI, Y. & SOLLER, M. 1990: Power of daughter and granddaughter designs for determining linkage between marker locus and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**:2525-2537.

WILCOX, M. D. 1983: Inbreeding depression and genetic variances estimated from self- and cross-pollinated families of *Pinus radiata*. Silvae Genetica **32**: 89-96.

WILLIAMS, C. G. & NEALE, D. B.: Conifer wood quality and marker-assisted selection: A case study. *Canadian Journal of Forest Research* **22**:1009-1017.

WILLIAMS, C. G. & SAVOLAINEN, O. 1996: Inbreeding depression in conifers: Implications for breeding. *Forest Science* **42**: 102-117.

YADZANI, R., MUONA, O., RUDIN, D. & SZMIDT, A. E. 1985: Genetic structure of *a Pinus sylvestris* L. seed tree stand and naturally regenerated under-story. *Forest Science* **31**: 430-436.

Chapter 4

MULTIPLE MARKER MAPPING OF WOOD DENSITY LOCI IN AN OUTBRED PEDIGREE OF RADIATA PINE

S. Kumar, R. J. Spelman, D. J. Garrick, T. E. Richardson,
M. Lausberg & P. L. Wilcox

THEORETICAL AND APPLIED GENETICS (2000) 100: 926-933

ABSTRACT

The objective of this study was to determine the genetic location and effects of genomic regions controlling wood density at three stages i.e., rings corresponding to ages 1-5 (WD1_5), rings corresponding to ages 6-10 (WD6_10) and outer wood density (WD14) in a full-sib pedigree (850.055 x 850.096) of *Pinus radiata*. The number of offspring measured at these three stages were 80, 93 and 93, respectively. Only a single linkage group of the parent 850.55 was considered for mapping quantitative trait loci (QTLs). A multiple marker least-squares approach was used for mapping QTLs for each of the three traits, using a single-QTL model. Logistic regression was used for multiple-trait QTL mapping. Critical values for test-statistic were calculated empirically by 'shuffling' the data. A putative QTL with large effect on WD1_5 appears to be segregating at 73cM position (experimentwise $P < 0.01$). The width of the 95% bootstrap confidence interval for this putative QTL was 40 cM (i.e. 56-96 cM). The effect of this QTL on the expression of wood density at later stages was diminished. From multiple-trait analysis, two marker locations (at 66 cM and 91 cM) were found to be significantly associated (experimentwise $P < 0.05$) with the expression of wood density at different ages. These results are encouraging for the application of marker information to early selection in order to increase juvenile wood density, although the putative QTLs detected in this study need to be verified in an independent population.

Key words: Quantitative trait loci (QTLs), *Pinus radiata*, wood density, linkage group

INTRODUCTION

Many traits of economic importance in plants and animals are of a quantitative nature. That is, the observed phenotypes are continuously distributed and reflect the action of many quantitative trait loci (QTL) together with environmental effects. The availability of genetic markers has allowed experimental studies in a number of species to explore the nature and location of some of these QTLs (DE KONING *et al.* 1998). These studies will provide insight into the control of these economically important traits, and may enhance breeding programmes through opportunities for marker-assisted selection.

QTL mapping studies in crop plants have been performed using segregating populations derived from crosses between inbred lines (e.g., TANKSLEY *et al.* 1982;

EDWARDS *et al.* 1987). Such populations are not available in trees and will be difficult to obtain for many species owing to high genetic load and long generation times. It is common to use full-sib families as mapping populations in QTL studies in outbred forest trees (e.g., PLOMION *et al.* 1996; KNOTT *et al.* 1997; EMEBIRI *et al.* 1998). Linkage analysis in a pedigree formed from crosses between two unrelated highly heterozygous trees is complex, as up to four alleles might be segregating at a locus. Tracking the inheritance of multiple alleles at QTL in an outbred pedigree necessitates the use of codominant multiallelic markers. However, for radiata pine there are presently insufficient multiallelic markers to construct a dense linkage map.

GRATTAPAGLIA & SEDEROFF (1994) adopted a 'two-way pseudo-testcross' approach with RAPD (random amplified polymorphic DNA) markers to construct linkage maps for each parent of a full-sib family of *Eucalyptus.* RAPD markers, in pseudo-testcross mapping configuration, are those which are in a heterozygous state in one parent and homozygous in the other, or vice-versa; therefore, separate sets of linkage data are obtained for each parent. As one of the parents of a full-sib pedigree (in pseudo-testcross mapping strategy) is non-informative at the RAPD marker loci (which means no information on QTL/marker linkages can come from non-informative parent), the inheritance of the gametes from the heterozygous parent can be analysed assuming a half-sib design. The advantage with this approach over the conventional half-sib design (where the marker information is usually available only on one parent) is that at each marker locus we know unequivocally which parent allele each offspring has inherited.

It is not uncommon to use the information from one marker at a time to analyse the marker-trait data using analysis of variance (ANOVA) techniques (WELLER *et al.* 1990; GROOVER *et al.* 1994). HALEY *et al.* (1994) showed bias in the estimated position of the QTLs can be reduced by including all markers in a linkage group. KNOTT *et al.* (1996), using a simulation study, showed that the multiple-marker technique has a higher power of detecting a putative QTL as compared to single-marker ANOVA method.

The present study demonstrates the use of the multiple marker techniques for the analysis of data from a pseudo-testcross mapping design, in a two-generation pedigree of radiata pine. A stepwise approach for the QTL analysis is presented here. Initially, we used exploratory analyses (VISSCHER & HALEY 1996; KNOTT *et al.* 1997; DE

KONING *et al.* 1998), in which trait scores are regressed onto selected marker information in an attempt to determine whether the inheritance of the trait (for individual linkage group) is compatible with an oligogenic model (several QTLs) or whether a small region of the linkage group (compatible with a single QTL) is important. This was followed by a more conventional search of the genome using a least-squares analysis (KNOTT *et al.* 1996). CHURCHILL & DOERGE's (1994) empirical method was used to obtain significance thresholds. Multiple-trait QTL mapping was also undertaken using logistic regression (e.g., HENSHALL & GODDARD 1999).

MATERIALS AND METHODS

Mapping population

Full-sib progeny of parent trees 850.055 and 850.096 of *Pinus radiata* were used to detect putative QTLs. This family was chosen because the parents displayed favourable general combining abilities, for various economic traits, in different mating designs. Tree 850.055 is widely used in seed orchards to produce seed for commercial plantations. The QTL detection population consisted of 80-93 trees (for different traits) from a large block Genetic Gain Trial planted in 1978 in Kaingaroa Forest in New Zealand that had not been subjected to any silvicultural selection.

Framework Linkage Map construction

Different dominant and codominant marker systems (RAPD, AFLP and SSR) were used to generate 430 polymorphic loci segregating in this pedigree. The linkage map of 850.055 consisted of 126 framework markers in 21 linkage groups, and covered 1540.2 cM, while the 850.096 map contained 101 markers in 26 linkage groups covering 1223.0 cM (P. L. WILCOX *et al.* unpublished). Markers were ordered on the map with a 1000:1 support. The genetic distances (cM) were obtained using the Kosambi map function. All of these framework markers were genotyped on 93 randomly selected individuals of this family. Preliminary analysis (P. L. Wilcox et al., unpublished) using single-marker ANOVA found marker-trait association on linkage group *three*. Thus, in this study, we considered markers only on linkage group *three* in the framework linkage map of the parent 850.055 for mapping the putative QTLs using the multiple marker method. The distribution of markers on this linkage group is shown in Table 1.

Table 1. Distribution of markers on linkage group *three*[a].

No.	Marker	Position on linkage group (cM)
1	A93_b3	0
2	A56_A	15
3	A184_b3	25*
4	RAPD_59	48*
5	RAPD_38	66*
6	A329_c3	83
7	A113_a1	91*
8	A297_b2	99
9	RAPD_270	106
10	A338_a1	115*
11	A219_b2	123
12	A140_c2	130
13	RAPD_192	140*
14	A47_c	153
15	RAPD_209	170*
16	A72_A	187

*denotes the markers selected for exploratory analysis.

[a] All markers were in a heterozygous state in parent 850.055 and homozygous null in the other parent (850.096).

Phenotypic data

Cores were extracted at age 14 and, for each individual tree, outer wood density (termed as WD14 in this study) was obtained from two samples of outer 50 mm of increment cores taken from opposite sides of the tree. To explore the potential for the use of core wood from young trees in QTL detection experiments, one 5 mm bark-to-pith core was taken from each tree in 1997. The wood density from these cores was measured using x-ray densitometry (COWN & CLEMENT 1983). Area weighted wood densities were obtained for rings corresponding to ages 1-5 years and ages 6-10 years. These two area-weighted wood densities were considered as two different traits and the symbols for them in rest of the text will be WD1_5 and WD6_10, respectively. Thus, there were three traits (WD1_5, WD6_10 and WD14) measured on each offspring.

Statistical methods

Exploratory analysis on multiple markers

An exploratory analysis of the linkage group was undertaken to determine whether the chromosomal region under study is associated with variation in the recorded traits (VISSCHER & HALEY 1996). To undertake the analysis, first, the locations of informative and evenly spaced markers are selected from those available in the data set. If too many markers are selected the analysis will take up a significant proportion of degrees of freedom. Information from closely related markers is highly correlated (DE KONING *et al.* 1998) and theoretical studies (DEKKERS & DENTINE 1991; VISSCHER 1996) indicate that markers spaced every 25 cM or so should explain most of the variation on a chromosome.

For exploratory analysis, data for each trait are regressed on selected marker positions. For a given position, the conditional probabilities of the offspring inheriting the first gamete of the parent provide an independent variable (as the probabilities of two parental alleles sum to unity) on which the trait values can be regressed (KNOTT *et al.* 1996). At the position of an informative marker we know which parental allele each progeny has inherited, so the probability of inheriting one allele will be unity and probability of inheriting the other will be zero. In pseudo-testcross mapping strategy, one parent is heterozygous at all marker loci but some faint RAPD bands are difficult to classify as presence or absence in some individuals. If a marker was not informative in a particular individual, it was replaced by the 'virtual' marker probability calculated for that position based on the nearest markers (e.g., KNOTT *et al.* 1996; DE KONING *et al.* 1998). On the assumption that the parental gamete reconstruction is correct, and for each offspring the allele inherited from the parent is known unequivocally, these probabilities are exactly the same as for a backcross situation (KNOTT *et al.* 1996).

The model for the exploratory analysis is:

$$Y_j = \mu + \sum_{k=1}^{n} b_k m_{jk} + e_j \qquad (1)$$

where

Y_j is the phenotypic value of offspring j

μ is the overall mean

b_k is effect of parental allele for marker k

m_{jk} is the probability for offspring j of inheriting the parental allele of marker k

e_j is the residual effect for offspring j

Comparison of alternative genetic models

Oligogenic

Using equation (1), the presence of genetic variation associated with a linkage group was tested by fitting the regression of each progeny phenotype on the conditional probabilities at all selected marker locations (n) simultaneously. Under the null hypothesis of no genetic variation for the trait associated with the linkage group under study only a family mean is fitted.

Single region

If an effect of a linkage group or chromosome is significant, further analysis can be used to identify whether there is one or more regions within the linkage group affecting the trait. The regression on all selected marker locations is compared with the regression on every pair of adjacent markers. Where more than one important QTL affects the trait, there will be no single pair of adjacent markers that accounts for as much variance as do all markers jointly (DE KONING *et al.* 1998). If the oligogenic model is not a significant improvement over fitting the best single region, it can be concluded that most of the genetic variance associated with this linkage group is explained by this single region.

Single-trait QTL analysis

The least-squares multiple-marker method described in KNOTT *et al.* (1996) was employed in this study. Basically, the model used here is same as that given in equation (1) except that the summation is now over the number of QTLs included in the model (one QTL in this study). At each 1-cM position, the values of phenotypes from progeny are regressed on the conditional probabilities.

Multiple-trait QTL analysis

The three traits measured (WD1_5, WD6_10 and WD14) on each individual can be considered as different states of the same trait. If each trait is analysed separately, it cannot be deduced whether these effects are due to one locus with correlated effects on these traits, or to several loci each affecting a different trait. The genetic correlations (based on an open-pollinated progeny test) among these juvenile and mature wood density traits were more than 0.75 (BANNISTER & VINE 1981). Multiple-trait QTL mapping could help in understanding a QTL's part in the genetic covariance structure of economically important traits. In this study, the logistic regression method proposed by HENSHALL & GODDARD (1999) was applied for the mapping of multiple-trait QTLs using each marker in turn. The parental allele (coded as 0 and 1) inherited by an individual, at a marker location, becomes the dependent variable and the phenotypes for three traits are the independent variable. Multiple-trait QTL effects, assuming no recombination ($r = 0$), were obtained as (HENSHALL & GODDARD 1999):

$$\boldsymbol{A} = \Sigma\beta / (1 + \mathrm{SQRT}(\beta' \Sigma \beta + 1)) \qquad (2)$$

where $\boldsymbol{A}$ is the vector of half the effect of allele substitution, Σ is phenotypic covariance matrix estimated from the complete experimental data, and β is the vector of parameter estimates. Σ was obtained using a phenotypic coefficient of variation of 0.07 for all three traits (COWN *et al.* 1992). The estimates of phenotypic correlations among these traits were obtained from BANNISTER & VINE (1981).

Significance thresholds

For single-trait QTL analysis

For the genome scans a large number of correlated tests with mixture distributions are being performed and, hence, the standard F distribution cannot be used to obtain the significance thresholds. An empirical distribution is therefore required, in order to test for significance. Test-statistic critical values were calculated empirically from the permutation method described in CHURCHILL & DOERGE (1994). The permutation test was undertaken by repeatedly randomly shuffling the phenotypic data. The conditional probabilities (m_{jk}'s) that the phenotypes are regressed on were not shuffled. The experimentwise critical values, which account for the evaluation of marker-QTL association across the genome and also the three traits being analysed, were calculated from the distribution of test-statistics (see SPELMAN *et al.* 1996). Because only one

chromosome is being analysed, the experimentwise critical values were calculated using the approximation of standard Bonferroni correction as:

$$\alpha \approx \gamma / n \qquad (3)$$

where α is the nominal threshold level to ensure γ significance level over the n (= R^*T) independent tests. The R (= 1540.2 / 187.0 = 8.24) denotes the ratio of total map length to the length of the chromosome under study and T denotes the number of independent traits. The number of independent traits were determined by factor analysis (using SAS 1988) on a genetic correlation matrix for the three traits. It was calculated that two factors account for approximately 98% of the variation. Thus, this suggests that there are only two independent traits (i.e., T = 2). The Bonferroni correction factor was applied to all three traits. The 100(1 - α) percentile of the distribution of test-statistics provides the experimentwise significance threshold.

Following DE KONING *et al.* (1998), for exploratory analysis, α levels of 0.0006 (=0.01 / (8.24*2)) and .0030 (= 0.05 / (8.24*2)) would be required to obtain 0.01 and 0.05 genome-wide levels, respectively, and can be obtained from standard tables. The suggestive level of significance (LANDER & KRUGLYAK 1995), where one significant result is expected by chance in a genome analysis, can be obtained from the binomial distribution as:

$$P_{suggestive} = 1 / n \qquad (4)$$

Many suggestive linkages will subsequently prove to be incorrect, but they are nevertheless reported so they can be followed up in future studies (KNOTT *et al.* 1997).

For multiple-trait QTL analysis

The logistic regression was performed at each marker location, in turn. Test-statistic critical values were calculated empirically, at each marker location, from the permutation method that involved repeated shuffling of three quantitative trait values together. The experimentwise critical value may be obtained by first finding the maximum test statistic over selected marker locations (about 25 cM apart) for each of the shuffled analyses. These values are then ordered and their 100(1 - α) percentile will provide the experimentwise significance threshold.

RESULTS

Quantitative traits

The phenotypic distributions of three traits examined are shown in Figure 1. The traits WD6_10 and WD14 were approximately normally distributed as tested by the Kolmogorov-Smirnov distance statistic (P-value > 0.15). Only WD1_5 showed a significant departure from normality (P-value < 0.01). The data for this trait was log-transformed to improve normality. However, results from QTL analysis using transformed data did not differ from those with untransformed data. Therefore, only untransformed data was used similar to some other studies (e.g., PLOMION *et al.* 1996; EMIBIRI *et al.* 1998). Phenotypic correlations were estimated among traits within the family studied. Significant ($P \leq 0.001$) correlations were observed between WD1_5 and WD6_10 ($r = 0.66$), WD1_5 and WD14 ($r = 0.37$) and WD6_10 and WD14 ($r = 0.54$).

Comparison of genetic models

Seven out of sixteen, approximately equally spaced (about 25 cM apart) markers were selected for this exploratory analysis (Table 1). The comparison of alternative genetic models is shown in Table 2. There is significant evidence, at a genome-wide 0.05 level, for a genetic component for the traits WD1_5 and WD14. However, the evidence of a genetic effect for the trait WD6_10 is significant only at a suggestive level (Table 2). The oligogenic model, for the traits WD1_5 and WD6_10, is not a significant improvement over the single-region model but, at a suggestive level, is significantly better in case of the trait WD14.

Table 2. Comparison of alternative genetic models.

Traits	WD1_5	WD6_10	WD14
N	80	93	93
# markers	7	7	7
Oligogenic[a,*]	0.0006	0.0211	0.0021
Oligo vs single interval[b,*]	0.5802	0.3772	0.0476

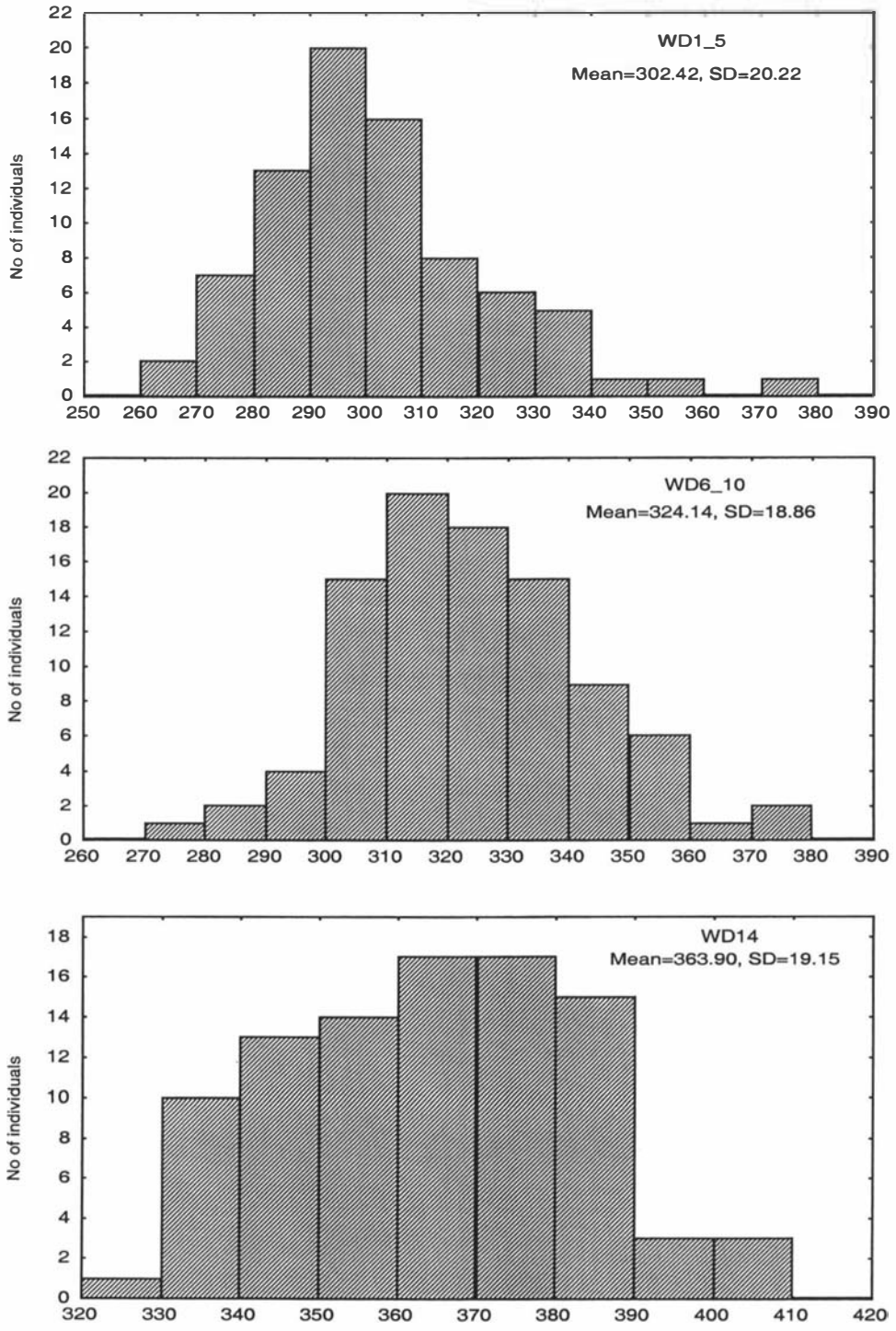
N denotes the sample size

[a] The probability of the F ratio for testing the oligogenic model versus a model with no genetic effect.

[b] The probability of the F ratio for testing the oligogenic versus best single-interval model.

*A probability level of 0.0006 (=0.01 / 8.24*2) and .0030 (= 0.05 / 8.24*2) would be required to obtain 0.01 and 0.05 genome-wide levels.

Figure 1. Phenotypic distributions for WD1_5, WD6_10 and WD14. Mean and standard deviation (SD) for each trait are also given.



Single-trait QTL analysis

Permutation test

The distribution of the test-statistic, by shuffling the trait value and fitting its regression on all marker positions in the linkage group, were obtained for each trait (Figure 2). These distributions are hardly distinguishable, which is also reflected from the critical values given in Table 3. The test-statistic distributions in Figure 2 account for repeated testing across the linkage group but do not account for repeated tests on the three correlated traits. Experimentwise threshold levels (Table 3) were calculated from the distributions in Figure 2 using the standard Bonferroni correction (Equation 3).

Figure 2. Approximate density function of test statistics for WD1_5, WD6_10 and WD14 derived from permutation test (50,000 shuffles).

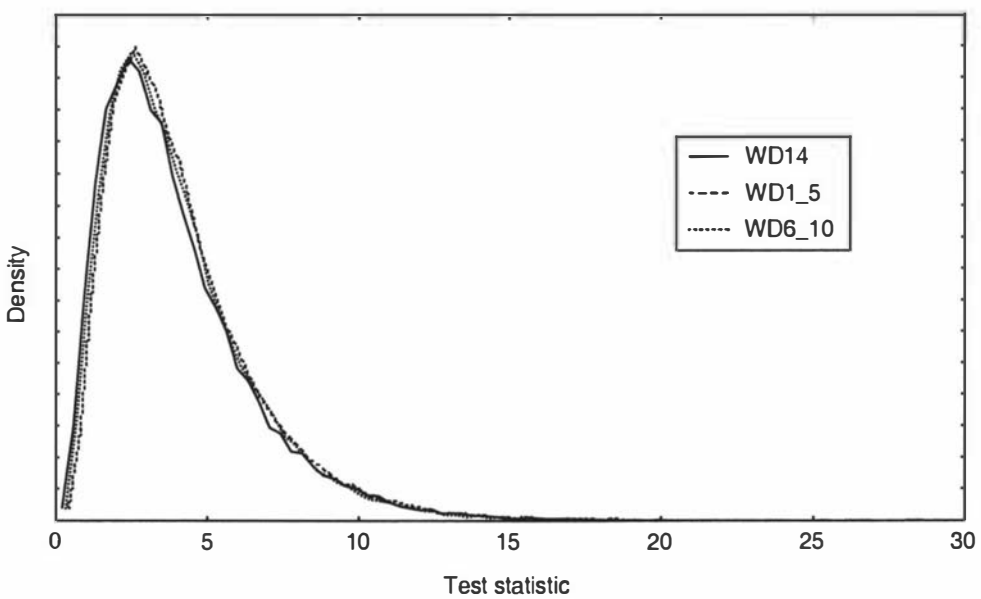


Table 3. Experimentwise threshold levels for the three traits (50,000 shuffles).

Threshold level	WD1_5	WD6_10	WD14
1%	18.69	18.53	18.99
5%	15.03	15.27	15.61
10%	13.45	13.61	14.02
15%	12.50	12.74	12.98

QTL mapping

For all three traits, the test-statistics calculated for a single-QTL model at each 1 cM are shown in Figure 3. The analysis revealed a putative QTL for WD1_5 positioned at 73 cM. The test-statistic was significant at the 0.01 genome-wide threshold level. The other traits, WD6_10 and WD14, showed an indication of a possible QTL at 57 cM and 135 cM, respectively (Figure 3). However, the test-statistic for WD6_10 and WD14 were only significant at the 0.18 and 0.15 genome-wide significance level, respectively. The estimated effects of allele substitution, at the most likely position of a QTL, were 23.17, 15.19 and 14.46 kg/m^3 for WD1_5, WD6_10 and WD14, respectively (Table 4).

Table 4. Estimated allele substitution effects and standard errors (S.E.), at the putative QTL, for different traits.

Trait	Position of the QTL	Effect (kg/m^3)	S.E.
WD1_5	73 cM	23.17	4.45
WD6_10	57 cM	15.19	4.33
WD14	135 cM	14.46	4.00

Multiple-trait QTL mapping

The approximate density function of the test-statistic, based on 5000 shuffles, is shown in Figure 4. Experimentwise threshold levels were calculated from the distribution in Figure 4 using the standard Bonferroni correction (equation 3). The test-statistics calculated at each marker location, using logistic regression, are shown in Figure 5. Specifically, there were two marker locations (RAPD_38 at 66 cM and A113_a1 at 91 cM) found to have association with putative QTLs at 0.05 genome-wide significance level. The effect of allele substitution (using equation 2) at marker RAPD_38 was 25.58, 23.12 and 17.89 kg/m^3 for WD1_5, WD6_10 and WD14 traits, respectively. Similarly, the substitution effect was estimated to be 22.56, 17.93 and 18.27 kg/m^3, at marker A113_a1, for the WD1_5, WD6_10 and WD14 traits, respectively.

Figure 3. Test statistics for different positions (at every 1cM) on chromosome *three* for different traits. Experimentwise threshold levels (1% and 15%) are also given.

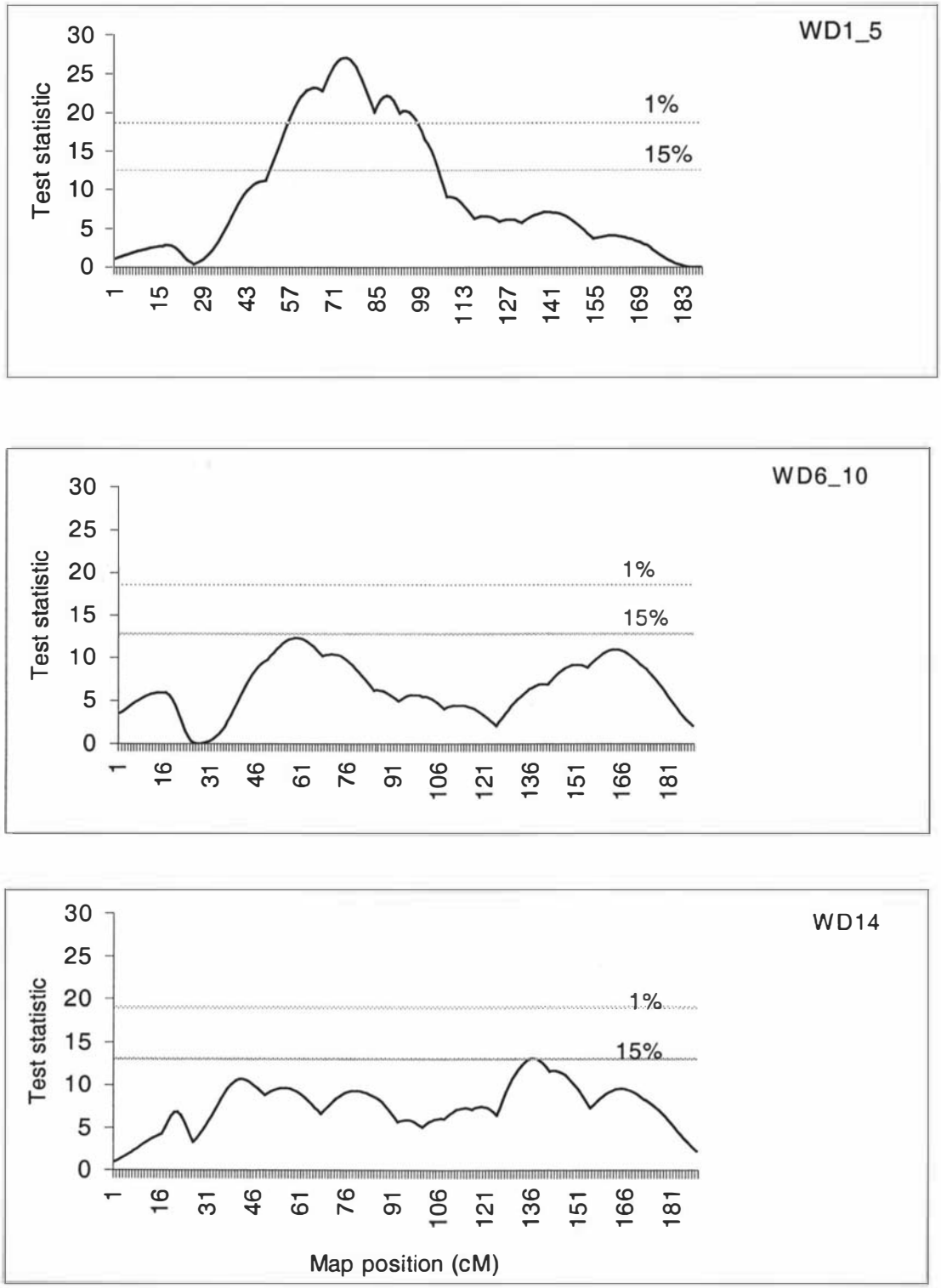


Figure 4. Approximate density function of the log-likelihood ratio test statistic (LRTS) derived from permutation test (5,000 shuffles).

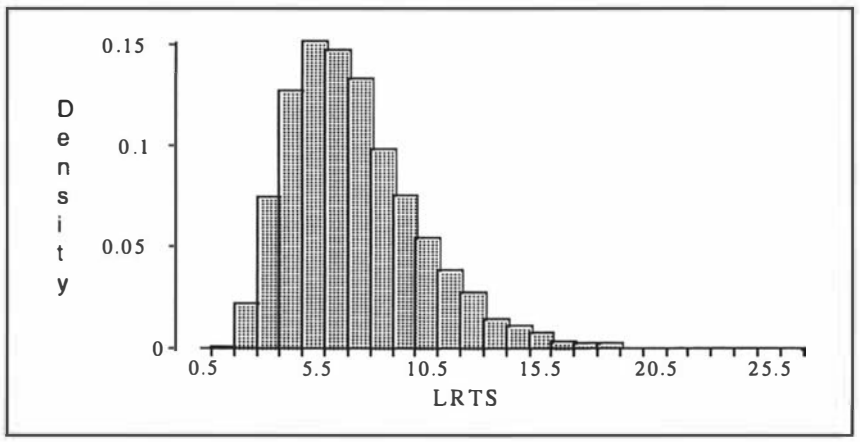
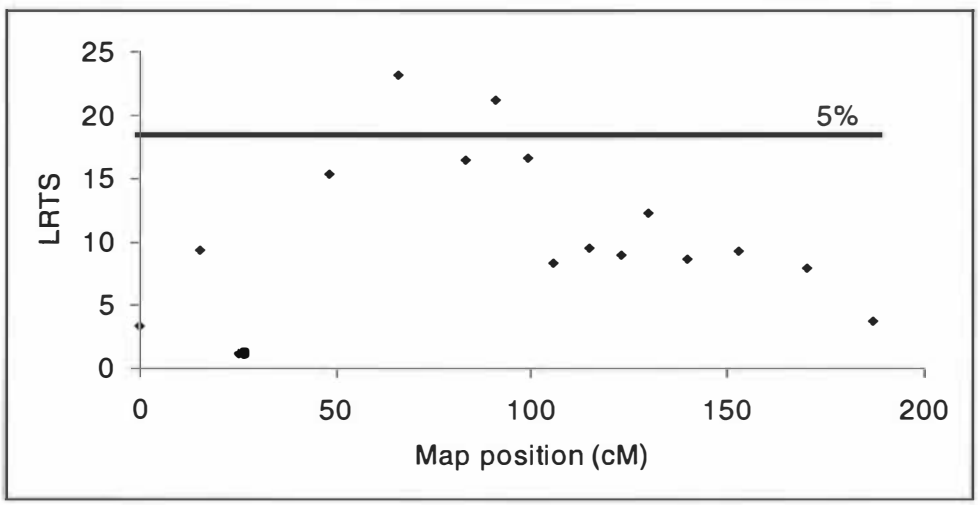


Figure 5. Log-likelihood ratio test statistic (LRTS) at marker positions. Experimentwise threshold level (5%) is also shown.



DISCUSSION

Comparison of alternate genetic models

Since the study by SAX (1923), and more recently with the development of molecular marker technologies, it has been demonstrated in plants and animals that a trait showing a continuous distribution could be under oligogenic control. In *Pinus taeda*, GROOVER *et al.* (1994) detected 5 QTLs for wood specific gravity that together explained 23% of the total variation. KNOTT *et al.* (1997) tested different genetic models for each linkage group in loblolly pine. They found different inheritance patterns for different linkage

groups. For most of the linkage groups contributing significantly to the genetic variation in trait, the single-region or oligogenic models were found to best explain the inheritance patterns. The present study revealed that for linkage group *three*, the inheritance patterns for different traits under study seem best explained by a single-region (one QTL) or small number of QTLs i.e., oligogenic (Table 2).

Single-trait QTL mapping

Using a single-QTL model, the highly significant evidence for a putative QTL was found only for the trait WD1_5 (Figure 3). The test for the presence of genetic variation associated with this linkage group, for trait WD1_5, was also significant at genome-wide 0.01 significance level and the oligogenic model failed acceptance in favour of a single-region model. There was little indication for the presence of genetic variation, for WD6_10, associated with this linkage group (Table 2) and thus there was no segment of this linkage group having an enormous effect on the trait variation. The putative QTL found for WD6_10 was significant at genome-wide significance level of about 0.18 (Figure 3). The improvement of the oligogenic model over the single-region model, for WD14, was only significant at a suggestive level and would have been rejected in favour of the single-region model. The QTL analysis revealed a putative QTL at 0.15 experimentwise threshold for the trait WD14.

The results showed that there is no strong statistical evidence of putative QTLs for the traits WD6_10 and WD14. The putative QTL found for WD1_5 had a considerable, but not significant, effect on the expression of wood density at later stages. As the genetic correlations among WD1_5, WD6_10 and WD14 are very high, BANNISTER & VINE (1981) concluded that in each tree the same genes were acting on wood density, and acting in much the same way, throughout the 15-year period. The results from the present study reflects that the same genes might be acting on wood density but the effect of gene substitution does not remain similar at all stages. For example, the effect of gene substitution at the putative QTL for WD1_5 (at 73 cM) was 23.17 kg/m^3 and it reduced to 13.32 and 12.83 kg/m^3 for WD6_10 and WD14 traits, respectively. This might be due to the weak environmental correlations among wood densities at different ages. As observed by SEARLE (1961), a phenotypic correlation less than its genetic counterpart, together with a small environmental correlation (as is the case among traits in this study), will occur where the genes governing two traits are similar but where the

environments pertaining to the expression of these traits have a low correlation. Environmental influence was suspected to be one of the causes for the differential expression of stem growth QTLs in radiata pine seedlings (EMEBIRI *et al.* 1998). These authors found that none of the putative QTL detected at any one stage were strongly expressed at all four stages of measurement. VERHAEGEN *et al.* (1997) found that no chromosomal region was consistently expressed across three ages for wood density in two species of *Eucalyptus*, though the analysis demonstrated the existence of a chromosomal segment being involved in the control of the trait across the period studied, independent of age. They also found that some QTLs were specific to a single stage.

Multiple-trait QTL mapping

Multiple-trait QTL mapping was undertaken at each marker location using a logistic regression method. Two marker locations (RAPD_38 at 66 cM and A113_a1 at 91 cM) were found significantly associated (at 0.05 experimentwise threshold level) with the expression of wood density at three stages i.e., WD1_5, WD6_10 and WD14 (Figure 5). Incidentally, these two markers are flanking the location of the putative QTL (at 73 cM) found for the trait WD1_5 and that had a considerable, though not significant, effect on other two traits as well. Using the bootstrap technique (see VISSCHER *et al.* 1996), we estimated the width of the 95% confidence interval (for the QTL found for WD1_5) to be 40 cM (i.e. 56-96cM).

Marker-Assisted Selection (MAS)

There is a high level of tree-to-tree genetic variation in *Pinus radiata* for various morphological traits, wood properties, disease resistance and many other traits for which actual data have been collected (BURDON 1992). This situation exists even in the progenies generated from inter mating parents of high breeding value. Thus, there is an opportunity to use marker-trait associations to increase genetic gain, per unit time, by selecting within the families that are used to establish production populations, and propagating for deployment only those individuals that have favourable marker genotypes. The first important step towards MAS is the detection and verification of QTL. Forest trees like radiata pine have long generation intervals and undergo various changes at morphological, anatomical and physiological levels during their life span. Instability of QTL expression over age has been reported in poplar for basal area

(BRADSHAW & STETTLER 1995), in maritime pine for juvenile growth (PLOMION *et al.* 1996; EMEBIRI *et al.* 1998) and in *Eucalyptus* for wood density (VERHAEGEN *et al.* 1997). However, the very high genetic correlations among the expression of wood density at different ages in radiata pine indicate that much the same genes are controlling the trait expression at different stages (BANNISTER & VINE 1981). The present study suggests that linkage group *three* in 850.055 contains such a locus. However, the effect of gene substitution would change at different stages. The putative QTL found for the trait WD1_5 was not significantly expressed at later stages. The putative QTLs for WD6_10 and WD14 were only significant at 15-20% genome-wide significance level. As the sample sizes used in this study (80-93) are small the power of QTL detection would be low (e.g., KUMAR *et al.* 2000) so that QTLs of low effect can not be detected. Also, with such a small sample size, there is a high probability that effect of allelic substitution is likely to be overestimated. Therefore, marker-trait associations found in this study need to be verified in an independent population. Nonetheless, this study indicates that early selection based on desired marker haplotypes might help increasing juvenile wood density.

ACKNOWLEDGEMENTS

We thank *Forest Research*, New Zealand, GEENZ Limited, FRST (Contract number CO4803) for their financial support. We also thank Drs SUE CARSON and ROWLAND BURDON for helpful discussion and comments on the manuscript, JOHN LEE for phenotypic data collection, GEOFF CORBETT and PAUL FISHER for genotyping work and C. CHENG for assistance with phenotypic data analysis.

REFERENCES

BANNISTER, M. H. & VINE, M. H. 1981: An early progeny trail in *Pinus radiata*. 4. Wood density. *N Z J For Sci* **11**:221-243.

BRADSHAW, H. D. & STETTLER, R. F. 1995: Molecular genetics of growth and development in Populus. IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. *Genetics* **139**:963-973.

BURDON, R. D. 1992: Genetic survey of *Pinus radiata*. 9: General discussion and implications for genetic management. *N Z J For Sci* **22**:274-298.

CHURCHILL, G. A. & DOERGE, R. W. 1994: Empirical threshold values for quantitative trait mapping. *Genetics* **138**:963-971.

COWN, D. J. & CLEMENT, B. C. 1983: A wood densitometer using direct scanning with x-rays. *Wood Sci Tech* **17**:91-99.

COWN, D. J., YOUNG, D. G. & BURDON, R. D. 1992: Variation in wood characteristics of 20-year-old half-sib families of *Pinus radiata*. *N Z J For Sci* **22**:63-76.

DE KONNING, D. J., VISSCHER, P. M., KNOTT, S. A. & HALEY, C. S. 1998: A strategy for QTL detection in half-sib populations. *Ani Sci* **67**:257-268.

DEKKERS, J. M. & DENTINE, M. R. 1991: Quantitative genetic variance associated with chromosomal markers in segregating populations. *Theor Appl Genet* **81**:212-220.

EDWARDS, M. D., STUBER, C. W. & WENDEL, J. F. 1987: Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* **116**:113-125.

EMEBIRI, L. C., DEVEY, M. E., MATHESON, A. C. & SLEE M. U. 1998: Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theor Appl Genet* **97**:1053-1061.

GRATTAPAGLIA, D. & SEDEROFF, R. R. 1994: Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* **137**:1121-1137.

GROOVER, A., DEVEY, M., FIDDLER, T., LEE. J., MEGRAW. R., MITCHEL-OLDS, T., SHERMAN, B., VUJCIC, S., WILLIAMS, C. & NEALE, D. 1994: Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**:1293-1300.

HALEY, C. S., KNOTT, S. A. & ELSEN, J. M. 1994: Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**:1195-1207.

HENSHALL, J. M. & GODDARD, M. E. 1999: Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression. *Genetics* **151**:885-894.

KNOTT, S. A., ELSEN, J. M. & HALEY, C. S. 1996: Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theor Appl Genet* **93**:71-80.

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. *Theor Appl Genet* **94**:810-820.

KUMAR, S., CARSON, S. D. & GARRICK, D. J. 2000: Detecting linkage between a fully-informative marker locus and a trait locus in outbred populations using analysis of variance. *Forest Genetics* **7**:47-56.

LANDER, E. & KRUGLYAK, L. 1995: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**:241-247.

PLOMION, C., DUREL, C.-E. & O'MALLEY, D. 1996: Genetic dissection of height in maritime pine seedlings raised under accelerated growth condition. *Theor Appl Genet* **93**:849-858.

SAS (1988) *SAS/STAT User's Guide, version 6.03.* SAS Institute, Cary, NC.

SAX, K. 1923: The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris. Genetics* **8**:552-560.

SEARLE, S. R. 1961: Phenotypic, genetic and environmental correlations. *Biometrics* **17**:474-480.

SPELMAN, R. J., COPPIETERS, W., KARIM, L., VAN ARENDONK, J. A. M. & BOVENHUIS, H. 1996: Quantitative trait loci analysis for five milk production traits on chromosome *six* in the Dutch Holstein-Friesian population. *Genetics* **144**:1799-1808.

TANKSLEY, S. D., MEDINA-FILHO, H. & RICK, C. M. 1982: Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity* **49**:11-25.

VERHAEGEN, D., PLOMION, C., GION, J.-M., POITEL, M., COSTA, P. & KREMER, A. 1997: Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers: 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor Appl Genet* **95**:597-608.

VISSCHER, P. M. 1996: Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. *J Heredity* **87**:136-138.

VISSCHER, P. M. & HALEY, C. S. 1996: Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. *Theor Appl Genet* **93**:691-702.

VISSCHER, P. M., THOMPSON, R. & HALEY, C. S. 1996: Confidence intervals in QTL mapping by bootstraping. *Genetics* **143**:1013-1020.

WELLER, J. I., KASHI, Y. & SOLLER, M. 1990: Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* **73**:2525-2537.

Chapter 5

GENETIC RESPONSE TO WITHIN-FAMILY SELECTION USING MARKERS IN SOME FOREST TREE BREEDING SCHEMES

S. Kumar and D. J. Garrick

(Submitted)

ABSTRACT

Marker assisted selection (MAS) provides an opportunity to increase the efficiency of within-family selection in forest tree breeding. Within-family MAS involves selection decisions first made on conventional breeding values, and QTL information used for within-family selection. In this study genetic response obtained by using MAS was compared with conventional methods for three options: '*full-sib family forestry*', '*clonal forestry*' and '*forwards selection for deployment*'. This comparison was undertaken using stochastic simulation for a locus that explained 20 % of the genetic variance. In '*full-sib family forestry*' scenario, markers were used to select genotypes (among juvenile individuals in a family) for vegetative propagation. Markers were used to pre-select genotypes for clonal testing in '*clonal forestry*' option. In case of '*forwards selection for deployment*' option, offspring that have favourable marker haplotype and a superior phenotype were selected from each family. The comparison between the MAS and the conventional strategy was evaluated in genetic terms based on comparison of the average genetic merit of the genotypes used for deployment in production plantations. The relative genetic gain (%) using MAS were found to be 6-8 % and 2-3 % higher compared to conventional strategy for '*full-sib family forestry*' and '*clonal forestry*' options, respectively. In case of '*forwards selection for deployment*' option, MAS was generally found to be providing higher genetic gain only when the heritability is low.

INTRODUCTION

Advances in molecular biology have presented tree breeders with the basis for a revolution in future improvement strategies, by providing novel opportunities to greatly enhance genetic gains. These gains can be additional to those from the continuation of conventional breeding strategies. Quantitative trait loci (QTL) for traits such as growth, wood quality and shoot phenology have been identified in conifers (e.g. GROOVER *et al.*, 1994; PLOMION *et al.*, 1996; KNOTT *et al.*, 1997; EMIBIRI *et al.*, 1998; KUMAR *et al.*, 2000). The major objectives of most, if not all, growth and wood-quality QTL studies is to identify QTL that can be utilised in marker assisted selection (MAS) breeding schemes. In forest trees, MAS can be beneficial in many ways. Selection efficiency can be improved from more accurate assessment of QTL genotype, and in time saving through early selection on seedling genotype for a trait expressed late in tree development.

Applying MAS in forest trees is very complex. Forest tree populations are characterised by high degrees of linkage equilibrium and, as a result, linkage phase relationship between markers and QTL will differ among individuals (STRAUSS *et al.*, 1992). Linkage equilibrium as well as the likelihood of QTL by genetic background interaction makes it difficult to use MAS for early selection, one of its most proclaimed benefits (TAUER *et al.*, 1992; WILLIAMS & NEALE, 1992). It will be necessary to screen for QTL in each full-sib family because of recombination and variation in background. Thus, phase relationships would have to be established for each parent tree. Marker alleles that are associated with positive QTL alleles in one family may be unassociated with QTL or associated with negative QTL alleles, in other families. This precludes the use of MAS for family selection, and requires that MAS for within-family selection be customised for each family. KERR *et al.* (1996), while evaluating MAS in *Eucalyptus* breeding programs assumed that marker-QTL associations were stable across a variety of genetic backgrounds. However, there is still not enough evidence of existence of linkage disequilibrium at population level in pine species. Within-family selections incorporating marker information is one practical option for implementation of MAS for tree breeding schemes in the immediate future.

Several strategies have been suggested to capture the gain from using within-family MAS. Within-family phenotypic selection is an important component of advanced-generation breeding plans in forest trees despite the poor response compared with family selection (e.g. van BUIJTENEN & BURDON, 1990; COTTERILL, 1986). Genetic markers can be used to increase the efficiency of phenotypic selection for low-heritability traits (LANDE & THOMPSON, 1990). O'MALLEY & MCKEAND (1994) suggested that MAS could be practised simultaneously with phenotypic selection in elite families to increase the efficiency of selecting individuals for production populations or for breeding in the next generation. Marker-trait associations could be utilised directly to increase genetic gain in production forests by selecting within the families that are used to establish production plantations and propagating only those genotypes that have favourable marker haplotypes for deployment. Further, increased genetic gain might also be achieved in concert with clonal forestry programs, where candidates for clonal testing are selected based upon favourable marker haplotypes.

Despite the feasibility and practicality of within-family MAS, there are few available reports that quantify likely gains from the alternate methods of applying within-family MAS in forest trees. In this study, the genetic gains obtained from utilising MAS for three different selection and deployment options available were compared for radiata pine. These three options are: 'family forestry', 'clonal forestry', and 'forward selection for deployment'.

SIMULATION MODEL

Population structure

A simple scheme for radiata pine improvement was modelled using stochastic simulation. A base population of 2000 parent trees was simulated. An open-pollinated progeny test, with 15 offspring of each base-population parent, was simulated to select 20 top parents. A seed orchard comprises these 20 backwards-selected parents. Also, these 20 parents were mated in a single-pair mating design to create 10 large full-sib family blocks (200 or 400 offspring per family). It is also assumed that the parents are kept in the seed orchard for supplying seed for commercial plantations for a long time (say 10-15 years).

It was assumed that a QTL and a linked marker had been identified via prior research. Marker and QTL genotypes were simulated for all trees in the base population. The QTL was assumed to be biallelic with equal allele frequency while the marker locus had four alleles with equal frequency. The linkage phase in the base population was known in order to simulate transmission of haplotype. For the simplicity, the recombination rate between marker and the QTL was assumed to be zero.

The base-population polygenic effects (u_i) were sampled from a normal distribution $N(0, V_a)$, where V_a is the polygenic variance which is additive genetic variance minus the variance due to the QTL (V_q) . Completely additive gene action was assumed and thus, the total genetic variance is same as the additive genetic variance. For a biallelic QTL with equal allele frequencies, the value of 'a' (which is half the difference between the genotypic values of the two QTL homozygotes) was calculated as the square root of $2V_q$ (see FALCONER 1989). The QTL component (V_q) was set to explain 20 % of the genetic variance. The effects of the QTL genotypes are assumed to be fixed

and thus same for all base parents. Residual (i.e. environmental) effects were sampled from normal distribution $N(0, V_e)$, where V_e is the residual variance, and added to the polygenic and QTL effects to quantify the phenotypic 'observations'. Different trait heritabilities (i.e. 0.25 and 0.75) were used.

Polygenic effects for the open-pollinated progeny of the base parents were generated from a normal distribution $N(\frac{1}{2} u_f, \frac{3}{4} V_a)$, where u_f is the polygenic effect of the female parent. The QTL effect and environmental effect were added to complete the phenotype of progeny. Base-parents were ranked based on the means of their open-pollinated progenies. The 20 top ranked parents were then mated in a single-pair mating design. The polygenic effect for the full-sib offspring was sampled from a normal distribution $N(\frac{1}{2} u_f + \frac{1}{2} u_m, \frac{1}{2} V_a)$, where u_m is the polygenic effect of the male parent. The QTL effect and environmental effect were added to the polygenic effect to determine the phenotype of each full-sib offspring.

Evaluation of MAS schemes

It was assumed that a QTL and a linked marker had been identified via prior research. Each family must be examined separately to determine whether the QTL and marker alleles are segregating, the linkage state of the marker alleles with respect to the QTL, and to verify that any measurable QTL effect exists in the particular genetic background. The required significance threshold that is used to identify whether the parents are segregating for the pre-identified QTL was held at 0.05. Large full-sib family blocks (with 200 or 400 offspring per family) were used to determine whether the seed orchard parents were heterozygous for the pre-identified QTL. The following are the three selection and deployment options (Figure 1) where MAS could be utilised. For each option, a MAS strategy was compared with a strategy that ignored any knowledge of segregating QTL (termed as control):

Option 1 – Full-sib family forestry

In the conventional 'full-sib family forestry' scenario, the control strategy (full-sib deployment) involved crossing the 20 top-ranked (seed-orchard) parents for deployment in commercial plantations as full-sib families.

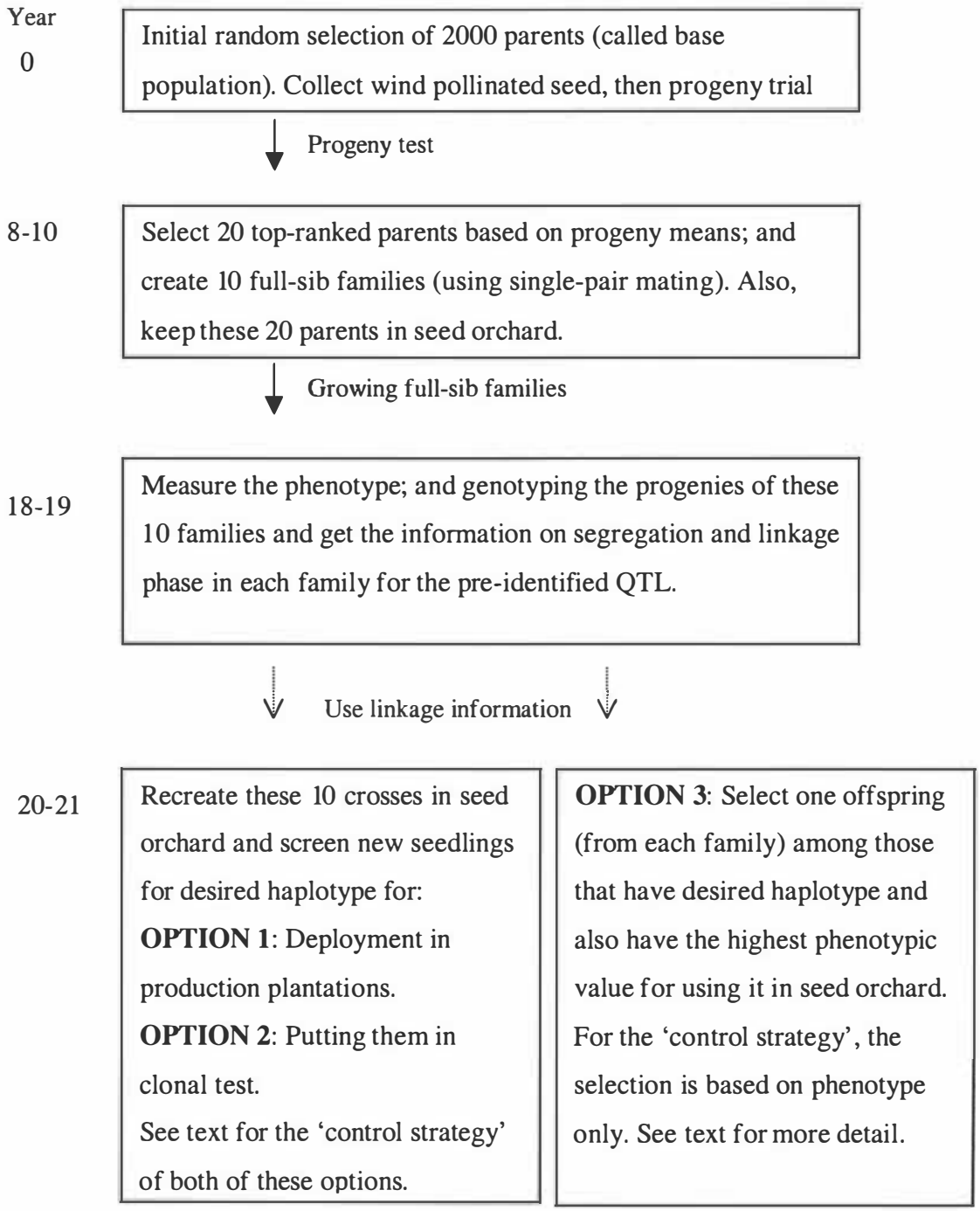
For the MAS strategy (*family MAS*), the information about the linkage phase and any measurable QTL effect for the pre-identified QTL needed to be examined in each of the 10 full-sib families. This step was undertaken in the large full-sib family blocks where the same 10 full-sib families are grown. Each of the 10 full-sib families were evaluated for pre-identified QTL by genotyping their progenies in the regions of interest. Offspring are grouped, depending on the marker haplotype they received, and the mean phenotype of different haplotype groups are compared to determine whether parents are heterozygous for the pre-identified QTL. The time required for adequate expression of growth traits is about 8 years in radiata pine. After obtaining the information on the pre-identified QTL for each family, the same crosses were recreated in seed orchard to produce new seeds and MAS was carried out by selecting 10 young seedlings from each family and then deploying these 100 genotypes, using vegetative propagation, directly in the forest.

The comparison between the MAS and the control strategy was evaluated in genetic terms based on comparison of the average genetic merit of the genotypes used for deployment in production plantations. In the control scenario, the average genetic merit is equivalent to the average breeding values of the 20 seed orchard parents. In the case of the MAS strategy, the genetic values of the 100 selected genotypes were averaged. The relative gain (%) using MAS was calculated as:

$$\text{Relative gain (\%)} = \frac{\text{mean genetic merit (MAS)} - \text{mean genetic merit (control)}}{\text{mean genetic merit (control)}} * 100$$

One hundred simulations were undertaken for each scenario. The average of relative gain (%) over the 100 replicates, standard deviation of relative gain and the proportion of replicates where MAS exceeded control gain are reported in this study.

Figure 1. Different options for applying MAS in tree breeding schemes.



Option 2- Clonal forestry

In the control strategy (*clonal testing*), the top 20 top-ranked (seed orchard) parents are crossed and either 30 or 60 young seedlings were selected randomly from each of the 10 full-sib families and were tested in field trials as clones. It was assumed that there are enough ramets to precisely estimate the genetic value of a clone. Ten unrelated clones (top ranked in each pair-cross) were selected for deployment in the commercial plantations.

For the MAS strategy (*MAS, then clonal testing*), the information as to whether the QTL and marker alleles were segregating, and the linkage phase of the marker alleles with respect to the QTL, was ascertained for the pre-identified QTL as explained in '*family MAS*' scenario for each of the 10 families. After obtaining the information (which would take about 8 years) on the pre-identified QTL for each family, the same crosses were recreated in seed orchard to produce new seeds and MAS was initially carried out for selecting either 30 or 60 genotypes with the most favourable marker haplotype from each full-sib family. These selected genotypes (total 300 or 600) were field-tested as clones and the 10 best unrelated clones were selected for deployment. The MAS breeding scheme and the control were evaluated in genetic terms based on comparison of the average genetic merit of the clones selected for deployment in production plantations.

Option 3- Forward selection for deployment

In the control strategy for this option, within-family phenotypic selection was carried out within the 10 full-sib families (obtained by crossing the 20 top-ranked parents) grown in large family blocks of either 200 or 400 offspring per family. The one best individual from each full-sib family was selected based on the phenotype alone. These 'forwards'-selected offspring were used in the seed orchard for commercial seed production. Selection of individuals to become the parents of the next generation breeding population would be done in a different way (say, two individuals from each full-sib family in order to maintain a constant census number of selected parents each generation). However, the scope of this scenario is limited to the selection of individuals for production purposes only.

To apply MAS in this scheme (*MAS plus phenotypic selection*), progenies of each of the 10 full-sib families were evaluated for information about linkage phase and QTL effect for the pre-identified QTL by genotyping the same trial as utilised in the control strategy. When a family is determined to be heterozygous for the QTL, only the offspring that has highest phenotypic value among those that received the most favourable haplotype is used in the seed orchard for commercial seed production. If any family is not segregating for the pre-identified QTL then the within-family selection is purely based on phenotype, just as in the control strategy. In this option, the time required for the application of MAS would be similar to that incurred in the control strategy. The only additional time in case of MAS would be in obtaining the marker genotypes of all offspring in each family. The MAS scheme and the control were evaluated in genetic terms based on comparison of the average genetic merit of the individuals selected for the production population.

RESULTS

Stochastic simulation was used to compare the genetic gain from some MAS strategies with that from conventional methods for three selection and deployment options. The relative gains (%) from MAS using 'full-sib family forestry' are shown in Table 1. In this scenario, young seedlings are selected based on the desired haplotype from each full-sib family and are deployed in forests using vegetative propagation. The results are shown for trait heritabilities of 0.25 and 0.75. Two different family sizes were used to evaluate each of the 10 full-sib families for the pre-identified QTL to determine the linkage phase and any measurable effect in each family. Slightly higher relative gains were obtained when larger family sizes were used to determine the linkage phase in each family. The magnitude of the relative gain using MAS varied from approximately 6 to 8 % for different trait heritabilites. Table 1 also shows the fraction of the replicates (i.e. number out of 100) where MAS was found to be providing higher genetic gain as compared to the control. In a few replicates the MAS strategy was found to have less gain compared to conventional full-sib deployment.

The genetic gains achieved from prior MAS of clones that go into the trial are shown in Table 2. The results are shown for different heritabilites (0.25 or 0.75) and the number of clones tested (30 or 60) from each of the 10 full-sib families. The relative genetic gains (%) captured from MAS were 2-3 % higher than the conventional approach of

'*clonal testing*'. In 84 to 93 % of replicates, the genetic gain achieved using MAS was found to be higher than that from conventional approach (Table 2).

Table 1. Average relative gain (%) from MAS using 'full-sib family deployment'. Figures in parenthesis are the corresponding standard deviations (based on 100 replications). 'Frequency' represents the number of times (out of 100) the MAS gave higher genetic gain as compared to the control.

Heritability	Family size	Relative gain (%) ± (SD)	Frequency
0.25	200	5.77 (4.53)	90
	400	7.99 (4.17)	100
0.75	200	6.41 (3.09)	99
	400	6.81 (2.86)	100

Table 2. Average relative gain (%) from marker assisted within-family clonal selection. Figures in parenthesis are the corresponding standard deviations (based on 100 replications). 'Frequency' represents the number of times (out of 100) the MAS gave higher genetic gain as compared to the control.

Heritability	# clones / family	Relative gain (%) ± (SD)	Frequency
0.25	30	2.37 (2.54)	84
	60	2.78 (2.11)	91
0.75	30	2.58 (1.97)	91
	60	2.65 (2.04)	93

MAS was also applied for forwards selection of individuals for use in the seed orchard (Table 3). From each of the 10 full-sib families that are planted in large family blocks, one individual that has the most favourable marker haplotype and a superior phenotype (among those that received the favourable haplotype) was selected as the seed-orchard parent. The relative gain achieved from using MAS compared to purely phenotypic selection is shown in Table 3. It shows that for low trait heritability (0.25), the MAS provided relative gain of order 1.10-1.48 %. For high heritability (0.75), conventional phenotypic selection was found to provide more genetic gain compared to the MAS.

Table 3. Average relative gain (%) from 'MAS plus phenotypic selection'. Figures in parenthesis are the corresponding standard deviations (based on 100 replications). 'Frequency' represents the number of times (out of 100) the MAS gave higher genetic gain as compared to the control (phenotypic selection).

Heritability	Family size	Relative gain (%) ± (SD)	Frequency
0.25	200	1.10 (4.57)	49
	400	1.48 (5.11)	63
0.75	200	-0.57 (2.05)	36
	400	-0.41 (2.05)	32

DISCUSSION

In this study, genetic gain obtained from using MAS was compared with conventional methods of selection and deployment. Three scenarios were considered: 'full-sib family forestry', 'clonal forestry' and 'forwards selection for deployment'. These three scenarios were not compared among each other. The comparison between MAS and the conventional method was made in genetic terms based on the comparison of average genetic merit of selected genotypes that were used for deployment in production plantations. However, in practice it will not be possible to know the genetic value of an individual unless it is progeny tested.

The size of the QTL (20 % of genetic variance or equivalent to 5 and 15 % of the phenotypic variance for trait heritability of 0.25 and 0.75, respectively) simulated in this study may seem intuitively large but QTL of this size have been reported in the literature. EMIBIRI et al. (1998) reported QTL explaining 6.8 % to 30.0 % of the variation in stem volume in radiata pine. PLOMION et al. (1996) reported QTL explaining 6.0 % to 20.4 % of the phenotypic variation in seedling height in maritime pine. M. M. SEWELL *et al.* (unpublished) reported 3 QTL explaining together 34 % of the phenotypic variance in wood specific gravity.

Full-sib family forestry

Full-sib deployment is one of the deployment strategies in commercial plantations. It is assumed that the seed orchard consists of 20 backwards-selected parents. The expected

genetic gain from this strategy would simply be equivalent to the average breeding values of the seed orchard parents involved in making the crosses. However, using molecular markers to select a few good seedlings from each cross would increase the genetic gain. Results given in Table 1 showed that MAS would provide about 6-8 % more genetic gain as compared to conventional way of full-sib family deployment.

For applying MAS, each family required to be examined separately for testing segregation and identification of the linkage phase for the pre-identified QTL. The time required for adequate expression of growth traits (like diameter and wood density) is about 8 years in radiata pine. This would delay the information on pre-identified QTL to be available and thus the increased gains using MAS would have to be evaluated in economic terms. Also, seed orchard parents would then have to be kept for longer time for supplying seed for commercial plantations. The other important consideration is the number of seedlings to be deployed from each family. Deploying 10 genotypes from each family would cover most of the risk of sampling error in their performance for the other traits not selected by MAS. However, this issue needs further investigation. Varying size of the full-sib family (200 or 400) used in the identification of whether parents were heterozygous for the pre-identified QTL, showed little effect on the increase in genetic level through MAS. Indeed, some studies in dairy cattle (MACKINNON & GEORGES, 1998; SPELMAN & GARRICK, 1998) found no effect of numbers of daughters used in the identification of whether a sire of sons was heterozygous, on the genetic level.

Clonal forestry

There is the possibility of increasing genetic gains by selecting genotypes for clonal testing based on molecular markers rather than random selection. Similar to '*family MAS*', the information on the segregation and the linkage phase, for the pre-identified QTL, comes from the larger full-sib family blocks. Thus, the additional time incurred in the application of MAS in this option would be about eight years. The increase in genetic gain by using MAS was around 2-3 % (Table 2). The additional genetic gains and the time incurred in using MAS need to be evaluated in economic terms. In both control and MAS strategies the same number of clones are being tested but the only difference is that in case of MAS the clones being tested are selected for the pre-identified QTL. Thus, the difference in genetic gain in these two options will depend

upon the size of the QTL. Varying size of the full-sib family (200 or 400) used in determining whether parents were heterozygous for the pre-identified QTL, showed almost no effect on the increase in genetic level through MAS (Table 2).

Forwards selection for deployment

In advanced-generations, forwards selection of individuals can be made either for selecting parents for next generation or for commercial seed production. Forwards selection from large full-sib family blocks is based on phenotypic values. Progeny testing candidates for within-family selection could double the generation interval and is seldom feasible. The accuracy of forwards selection could be increased by choosing offspring that have favourable QTL genotypes and a superior phenotype. The comparison of purely phenotypic selection with that using markers and phenotype (Table 3) showed that the latter provide higher average gain compared to the former for a trait heritability of 0.25. However, for high heritability (0.75) the phenotypic selection is more efficient compared to selection based on marker haplotype and phenotypic information. The required significance threshold that is used to identify whether the parents are segregating for the pre-identified QTL was kept at 0.05. Lower threshold levels (i.e. 0.10 or 0.20) will increase the chance of selecting a wrong haplotype and can reduce selection efficiency. However, increasing the threshold level, in this case, did not increase the average genetic level of the selected individuals (results not shown).

A major difference between control and MAS in this option arises in the selection intensity. In case of phenotypic selection, one best individual is selected out of the total available (say 200 or 400) in the family. However, for MAS the best individual is selected out of only those that received the desired haplotype. Also, as the tree gets older and the heritability of the trait is very high (like 0.75), phenotypic selection should be highly efficient as compared to low heritability (0.25). This might be the reason that average relative gain (Table 3) using MAS is higher than phenotypic selection only when trait heritability is 0.25. This issue needs to be further investigated.

It can be noticed from the results that the standard deviation (based on 100 replications) of the average relative gain (%) is quite high. Particularly, for '*Option 3*', the standard deviations are higher than the average relative gain (%). It indicates that MAS could

sometime come out worse than the conventional selection unless markers are explaining a high proportion of the genetic variance. Also, instead of one QTL explaining 20 % of the genetic variation (as assumed in this study) if there are say four QTLs each explaining 5 % of the genetic variance, then the relative gains would be less than that reported in this study. In this study, it was assumed that marker and QTL are closely linked and thus the results presented should be considered upper limits for the genetic gains possible with MAS.

ACKNOWLEDGEMENTS

I would like to thank Drs Sue Carson, Rowland Burdon, Richard Spelman and Randy Johnson for suggestions and comments in the preparation of this manuscript. Thanks to ***Forest Research***, New Zealand for financial support.

REFERENCES

COTTERILL, P. P. 1986: Genetic gains expected from alternative breeding strategies including simple low cost options. *Silvae Genetica* **35**:212-223.

EMEBIRI, L. C., DEVEY, M. E., MATHESON, A. C. & SLEE, M. U. 1998: Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theoretical and Applied Genetics* **97**:1053-1061.

FALCONER, D. S. 1989: *Introduction to Quantitative Genetics*, 3rd ed. Longman Scientific & Technical, Essex, England. 438 pp.

GROOVER, A., DEVEY, M., FIDDLER, T., LEE, J., MEGRAW, R., MITCHEL-OLDS, T., SHERMAN, B.,VUJCIC, S., WILLIAMS, C. & NEALE, D. 1994: Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**:1293-1300.

HALDANE, J. B. S. 1919: The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genetics* **2**:3-19.

KERR, R. J., JARVIS, S. R. & GODDARD, M. E. 1996: The use of genetic markers in tree breeding programs. Pp. 498-505 *in* M. J. DIETERS, A. C. MATHESON, D. G. NIKLES, HARWOOD, C. E. & MALKER, S. M. (eds). *Tree Improvement for Sustainable Tropical Forestry*. Proc. QFRI-IUFRO Conference, Caloundra, Queensland, Australia, 27 Oct – 1 Nov 1996.

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.

KUMAR, S., SPELMAN, R. J., GARRICK, D. J., RICHARDSON, T. E., LAUSBERG, M. & WILCOX, P. L. 2000: Multiple marker mapping of wood density loci in an outbred pedigree of radiata pine. *Theoretical & Applied Genetics* **100**:926-933.

LANDE, R. & THOMPSON, R. 1990: Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**:743-756.

MACKINNON, M. J. & GEORGES, M. A. J. 1998: Marker-assisted pre-selection of young dairy sires prior to progeny testing. *Livestock Production Science*, **54**:229-250.

O'MALLEY, D. M. & MCKEAND, S. E. 1994: Marker assisted selection for breeding value in forest trees. *Forest Genetics* **1**:207-218.

PLOMION, C., DUREL, C.-E. & O'MALLEY, D. M. 1996: Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. *Theoretical and Applied Genetics* **93**: 849-858.

SPELMAN, R. J. & GARRICK, D. J. 1998: Genetic and economic responses for within-family marker assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science* **81**:2942-2950.

STRAUSS, S. H., LANDE, R. & NAMKOONG, G. 1992: Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research* **22**:1050-1061.

TAUER, C. G., HALLGREN, S. W. & MARTIN, B. 1992: Using marker aided selection to improve tree growth response to abiotic stress. *Canadian Journal of Forest Research* **22**:1018-1030.

VAN BUIJTENEN, J. P & BURDON, R. D. 1990: Expected efficiencies of mating designs for advanced-generation selection. *Canadian Journal of Forest Research* **20**:1648-1663.

WILLIAMS, C. G. & NEALE, D. B. 1992: Conifer wood quality and marker-aided selection: a tool for the improvement of forest tree species. *Canadian Journal of Forest Research* **22**:1009-1017.

Chapter 6

IMPACT OF ERRORS OF ESTIMATING MULTIVARIATE GENETIC PARAMETERS ON SELECTION EFFICIENCY AND EFFICACY OF ALTERNATIVE METHODS OF OVERCOMING SUCH ERRORS

S. Kumar, R. D. Burdon & D. J. Garrick

FOREST GENETICS (2000) 7: 39-46

ABSTRACT

Selection index gives optimal index weights and maximises expected genetic gain when variance-covariance parameters are known exactly. However, in practice this is seldom the case; the parameters must be estimated. The present study was conducted to overcome the effects of sampling errors on selection efficiency. A method is proposed which consists of 'regressing' the estimated (least-squares) selection index coefficients ($\hat{\boldsymbol{b}}$) towards the relative economic values ($\boldsymbol{a}$), which are assumed to be known precisely, as: $\hat{\boldsymbol{b}}^{*} = \hat{\boldsymbol{b}}k + \boldsymbol{a}(1-k)$, for $0 \leq k \leq 1$. The efficiency of the proposed method along with some other index selection procedures (e.g., unmodified index selection, base index and "bending") was evaluated for 192 parameter and sample situations with 1000 replicates each, by Monte Carlo simulation. The highest gain was associated with $k < 1$. The mean improvement in % gain obtainable with optimal k (over $k = 1$) was 40.4, 17.7, 6.7 and 2.3 for sample size of 25, 50, 100 and 200 families, respectively. The optimum k-value increased with increase in sample size but decreased as the number of traits in the index increased. The relative efficiency of the proposed method was higher compared to other procedures. When the heritabilities of index-traits were low and their relative economic values were in opposite order to heritability, the relative efficiency of the proposed method was much higher.

Key words: sampling error, selection index, economic weights, index weights, efficiency.

INTRODUCTION

The theoretical basis of index selection was developed by SMITH (1936) and HAZEL (1943), and involves the indirect selection of an unobserved variable, H, by truncation selection of an observed variable, I, which is jointly distributed with H. The index (I) is a linear function of observations, which aims at ranking the population for aggregate genotype, i.e., H (KEMPTHORNE 1957, FALCONER 1989). Index weights can in principle be found by using a least-squares solution to minimise prediction error, or equivalently maximise the product-moment correlation between index values and aggregate genotype.

An index $I = \boldsymbol{b}'X$, where X is a vector of phenotypic deviates from fixed constants (for example, site means) that are assumed to be known, on p traits, is generally used in order to maximise the correlation with overall aggregate genotype $H = \boldsymbol{a}'\boldsymbol{g}$, where $\boldsymbol{a}$ is a vector of known economic weights and $\boldsymbol{g}$ is vector of true breeding values on the same p traits. Suppose $\boldsymbol{P}$ = var (X) and $\boldsymbol{G}$ = var (g) are the phenotypic and genetic covariance matrices, respectively, the optimum index is given by

$$\boldsymbol{b} = \boldsymbol{P}^{-1}\boldsymbol{G}\boldsymbol{a} \qquad (1)$$

The expected response per generation to selection is:

$$R = i\,(\boldsymbol{b}'\boldsymbol{P}\boldsymbol{b})^{1/2}, \qquad (2)$$

where i is the selection intensity. By using estimates of $\boldsymbol{P}$ and $\boldsymbol{G}$, namely $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{G}}$, the estimated index weights are:

$$\hat{\boldsymbol{b}} = \hat{\boldsymbol{P}}^{-1}\hat{\boldsymbol{G}}\,\boldsymbol{a}, \qquad (3)$$

and predicted genetic response is

$$\hat{R} = i\,(\hat{\boldsymbol{b}}'\,\hat{\boldsymbol{P}}\,\hat{\boldsymbol{b}})^{1/2}. \qquad (4)$$

The derivation of the selection index is based on the assumption that the population parameters such as heritability (h^2), genetic and phenotypic correlations, phenotypic standard deviations or alternatively genetic and phenotypic variance-covariance matrices, are known exactly. In practice, however, only estimates of these parameters are usually available for constructing the index, and such an index is less efficient than one computed from the true parameters. The effects of errors in the parameter estimates and the loss in efficiency, in relation to size of the sample used for estimation, have been considered by WILLIAMS (1962a, b), HARRIS (1964), and SALES & HILL (1976a, b). With multi-trait indices, it appears that rankings on index values will be more sensitive to errors in estimating genetic and phenotypic covariance matrices if the traits in the index are adversely correlated than if they are favourably correlated (BULMER 1985).

HAYES & HILL (1981) proposed a technique called 'bending' for modifying parameter estimates for multi-trait individual selection. If phenotypic ($\boldsymbol{P}$) and genetic ($\boldsymbol{G}$) variance- covariance matrices are estimated from between- and within-class covariance matrices, $\boldsymbol{B}$ and $\boldsymbol{W}$ respectively, in a one-way multivariate analysis of variance, then

according to the method the bent genetic and phenotypic covariance matrices, $\hat{G}^*$ and $\hat{P}^*$ are

$$\hat{G}^* = 4(B^* - W) / n, \qquad \hat{P}^* = \{B^* + (n\text{-}1)W\} / n$$

where

$B^* = (1 - \gamma)B + \bar{v}\gamma W$, where γ is the 'bending' factor and $\bar{v}$ denotes the average root of $W^{-1}B$. For $\gamma = 0$, $\hat{P}^{*-1}\hat{G}^* = \hat{P}^{-1}\hat{G}$ or $\hat{b}^* = \hat{b}$; and for $\gamma = 1$, $\hat{P}^{*-1}\hat{G}^* = I$ or $\hat{b}^* = a$. HAYES & HILL (1981) propose two alternative procedures: (i) if any roots of $\hat{P}^{-1}\hat{G}$ are negative, bend until the smallest root is zero; (ii) bend on the basis of the sample size alone.

ARNASON (1982) used 'bending' to predict the breeding values for multiple traits in a small, non-random-mating (horse) population. MEYER & HILL (1983) extended this bending procedure to the case when both individual and sib-information are available, allowing also for different subsets of traits being represented as characters in the selection criteria (in the index) and traits in the (economic) aggregate genotype, respectively. MEUWISSEN & KANIS (1988) used a bending procedure to make an inconsistent set of contrived population parameters (taken from several sources) consistent. A 'rounding procedure' was proposed by TAI (1989) to improve the efficiency of index selection, which involves performing canonical variate analysis on phenotypic and genetic variances of a group of traits estimated from a progeny test experiment.

The ridge regression technique (HOERLE & KENNARD 1970) was devised to circumvent the problem of an ill-conditioned covariance matrix of independent variables in multiple regression analysis, and has been used to develop a ridge selection index (SAXTON 1986, XU & MUIR 1989, VERRYN 1994). Using this procedure, the modified index weights can also be calculated as

$$\hat{b}^* = \left[\hat{P} + \delta \mathrm{diag}\left(\hat{P}\right)\right]^{-1}\hat{G}a, \quad \delta \geq 0$$

SAXTON (1986) applies "ridge regression" and "bending" to prediction in breeding. Bending performed better than the ridge procedures, and ridge procedures performed better than least-squares (LS) selection index (SAXTON 1986).

To circumvent the problem of choosing an optimal bending factor for a given sample situation, ESSL (1991) proposed the use of prior knowledge of the genetic parameters. Also, he suggested to use that bending factor which maximises the correlation between true and estimated aggregate genotype as

$$r_{IH} = \frac{a'G\hat{b}^*}{(a'Ga)^{0.5}\left(\hat{b}^{*'}\hat{P}\hat{b}^*\right)^{0.5}}$$

The vector of modified index weights, $\hat{b}^*$, is obtained by bending $\boldsymbol{B}$ towards $\boldsymbol{W}$ as suggested by HAYES and HILL (1981). Because the $\boldsymbol{G}$ matrix is unknown, he suggested to use its prior estimate. This new bending strategy of ESSL (1991) was found to be better than those two suggestions of HAYES and HILL (1981) while dealing with a given sample situation.

An alternative viewpoint, proposed in this paper is to give more weight to the economic information when there is doubt about the accuracy of the estimates of genetic parameters. The extreme approach is to use a so called 'base index' (WILLIAMS 1962a, b). An apparently unexplored refinement of this procedure is to regress the computed index towards the base index. The present study was aimed at deriving satisfactory index weights assuming good economic information. The efficiency of the proposed method was compared with the bending approaches of HAYES & HILL (1981) through Monte Carlo simulation.

MATERIALS AND METHODS

There is assumed to be a one-way classification with f groups or half-sib families each of size n, and p traits are recorded on each individual. It is further assumed that the observations are multivariate normally distributed with among- and within-group effects independent of each other. The multivariate analysis of variance table, in the notation of HAYES & HILL (1981), is as follows:

Source	df	SS	MS	E(MS)
Among groups	f - 1	$\boldsymbol{S_B}$	$\boldsymbol{B}$	$\Sigma + n\Psi$
Within groups	$f(n$-1)	$\boldsymbol{S_W}$	$\boldsymbol{W}$	Σ

The matrices of sums of squares and cross-products follow independent central Wishart distributions, $S_B \sim W_p [(f - 1), \Sigma + n \Psi]$ and $S_w \sim W_p [f(n - 1), \Sigma]$. An estimate of Σ is $\boldsymbol{W}$, whereas the estimate of $\Psi = 0.25\boldsymbol{G}$ is $(\boldsymbol{B} - \boldsymbol{W})/ n$. The dimension of each matrix is p x p and W_p signifies a Wishart distribution with p as number of variables. The matrix Σ is positive definite and Ψ is positive semi-definite.

It is assumed that 'mass selection' is practised. The expectation of response that is actually achieved when $\hat{\boldsymbol{b}}$ (equation 3) is used subsequently for making selection decisions in the population is

$$\hat{R}^a = i\, \hat{\boldsymbol{b}}' \boldsymbol{G}\boldsymbol{a} (\hat{\boldsymbol{b}}' \boldsymbol{P} \hat{\boldsymbol{b}})^{-1/2}. \qquad (5)$$

As the index coefficients in (3) are vulnerable to sampling errors in $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{G}}$, the expected gain values in (4 and 5) are themselves sensitive to these errors. The estimated index coefficients were regressed towards the relative economic values (REVs), assuming that the REVs are known precisely, as follows:

$$\hat{\boldsymbol{b}}^* = \hat{\boldsymbol{b}} k + \boldsymbol{a} (1-k),\ 0 \leq k \leq 1 \qquad (6)$$

where $\hat{\boldsymbol{b}}^*$ is the vector of modified index weights. Clearly, when $k = 1$, the selection is solely based on estimated LS index weights and when $k = 0$, the index is 'base index'. These modified index weights were used in place of $\hat{\boldsymbol{b}}$ in (5) to calculate the expected value of genetic response. In this way, some optimum value of k can be sought in order to maximise the value of expected gain. This strategy of finding the optimum k value is theoretically similar to that of finding optimum γ which maximises the correlation between true and aggregate genotype, proposed by ESSL (1991). The optimum value of k was considered to be that which gives the maximum average expected genetic gain (5).

The true achievable genetic response by using the optimum index weights, $\boldsymbol{b}$ (1), was calculated as

$$R^a = i\, \boldsymbol{b}' \boldsymbol{G}\boldsymbol{a} (\boldsymbol{b}' \boldsymbol{P}\boldsymbol{b})^{-1/2}, \qquad (7)$$

which is the maximum achievable gain from the known genetic and phenotypic covariance matrices.

Simulation study

Monte Carlo simulation was used to study the effect of sampling errors on the selection index coefficients. In general, the environmental correlation was assumed to be zero but a few sets of parameters were considered to study the effect of non-zero environmental correlations. The number of traits simulated in the index were 2, 3 or 4. Different sets of economic weights were considered, i.e. equal, in same rank order as heritabilities and in opposite rank order to heritabilities. The sampling intensity and the phenotypic variance of each trait were assumed to be unity. Different sample sizes considered were 25, 50, 100 and 200 half-sib families. The number of individuals per family was kept constant at 15. The sets of the assumed parameters used for simulation are given in Table 1. Some of these parameters are as used by ESSL (1991).

The parameter sets where non-zero environmental correlations were assumed are also given in Table 1. The properties of these real (assumed) parameter matrices were examined in terms of their eigenvalues and all eigenvalues of these matrices were positive. The number of replicate simulation runs for each scenario was 1,000. For each replicate run, the among- and within-group matrices of sums of squares and cross-products (for varying sample sizes) were sampled independently from Wishart distributions. The genetic and phenotypic variance-covariance matrices were estimated from the sampled among- and within-group matrices of sums of squares and cross-products. Different index selection procedures compared for their efficiency, were:

1. $k = 0$: Equivalent to the base index of WILLIAMS (1962a, b),
2. $k = 1$: Index based on unadjusted sample estimates,
3. kR : Stepwise procedure by progressively regressing the estimated index coefficients towards REVs using (6),
4. γR : Stepwise bending procedure by progressively bending $\boldsymbol{B}$ towards $\boldsymbol{W}$ (HAYES & HILL 1981).
5. γN : Bending procedure using fixed γ values (suggestion (ii) of HAYES & HILL 1981).
6. kN : regressing the estimated index coefficients towards REVs using fixed k values (based on sample size alone).

Table 1. Parameter sets (heritability, h^2, and genetic correlations, r_g) used in simulation. Figure in parenthesis, if any, are the corresponding environmental correlations.

Set No.	h_1^2	h_2^2	h_3^2	h_4^2	$r_{g(1,2)}$	$r_{g(1,3)}$	$r_{g(1,4)}$	$r_{g(2,3)}$	$r_{g(2,4)}$	$r_{g(3,4)}$	Economic Weights*
1	.20	.70			-.30						11 12 21
2	.15	.45			-.70						11 12 21
3	.25	.35			.40						11 12 21
4	.05	.25			-.50						11 12 21
5	.25	.30	.35		.10	-.30		-.20			111 123 321
6	.15	.30	.45		.30	-.50		-.40			111 123 321
7	.05	.30	.60		.40	-.70		-.60			111 123 321
8	.05	.15	.20		.30	-.50		-.40			111 123 321
9	.20	.30	.30	.40	.10	0	-.20	0	-.10	0	1111 1233 3211
10	.10	.25	.35	.50	.20	.10	-.50	0	-.40	-.20	1111 1233 3211
11	.05	.15	.30	.70	.40	.10	-.80	0	-.60	-.20	1111 1233 3211
12	.05	.10	.15	.20	.20	.10	-.50	0	-.40	-.20	1111 1233 3211
13	.25	.40	.60		.15 (.08)	.20 (.15)		-.35 (-.10)			111 123 321
14	.15	.30	.45		.50 (.30)	.30 (.10)		-.40 (-.20)			111 123 321
15	.38	.31	.53	.70	.15 (.15)	.62 (.32)	-.29 (-.14)	.28 (.22)	-.16 (-.01)	-.20 (-.01)	1111 2133 2311
16	.21	.14	.26	.33	.16 (.29)	.43 (.34)	.11 (.14)	.13 (.19)	.04 (.01)	-.08 (-.01)	1111 2133 2311

* The value of economic weights, for example 11, means equal weights are given to two traits

For the procedure γR the optimum γ value was chosen which maximises expression (5). This is theoretically similar to the proposal of ESSL (1991) to use that bending factor which maximises correlation between the true and the estimated aggregate genotype. It is however different from the first suggestion of Hayes & Hill which states that if any roots of $\hat{G}$ or ($\hat{P}^{-1}\hat{G}$) are negative, bend until the smallest root is zero. The ridge index selection procedure was also evaluated but it was found to be consistently inferior and thus it will not be discussed further.

For each sample run, the procedures, i.e. kR and γR were applied using the full range (0 to 1) of constants k and γ with increment values of 0.01. The expected genetic gain (5) was calculated for each sample for different values of k and γ. After that, these gain values were averaged over all samples (i.e. 1000). All comparisons were applied on the average values of (5). For the comparisons, i.e., kR and γR, the best 'fixed' value of k or γ was considered to be that which maximises average value of expression (5). The recommended γ values of HAYES & HILL (1981) apply to the case of 16 individuals per family. Although the individuals per family were kept at 15 in this study the fixed γ values from HAYES & HILL (1981) were used as such for calculations of procedure γN. The efficiency of all procedures investigated was judged by average expected gain (5) relative to an index with true index weights (7). A computer program to generate Wishart distributions was written in SAS/IML (1989) using the algorithm of ODELL & FEIVESON (1966), as explained in KENNEDY & GENTLE (1980).

RESULTS

The results include replicates where estimated genetic correlations fell outside the theoretical range -1 to 1 and also cases of negative heritability estimates, particularly when the number of families were small i.e. $f = 25$. However, the average estimates of these parameters were almost identical to the true parameters even with a sample size of 25 families with 15 individuals each. For a sample of 100 families, the parameter estimates, i.e. of heritability and genetic correlations, were found to fall consistently well within the theoretical bounds i.e., $0 \leq h^2 \leq 1$ and $-1 \leq r_g \leq 1$. Similar to HAYES & HILL (1981), the replicates where genetic parameters fell outside the parameter space were included in further calculations.

The plot of results for parameter set 8 (see Table 1) is shown in Figure 1. The number of families for this case are 50 and REVs of traits were in opposite rank order to the heritabilities. The Figure 1 consists of average (over 1000 replications) expected gain (equation 5) for two different procedures, i.e., kR and γR, over full range of k and γ. It can be seen from this figure that maximum average expected gains are 0.3018 (at k = 0.88) and 0.2947 (at γ = 0.56) for kR and γR procedures, respectively. The optimal genetic gain (7) for this case was 0.3527. Thus, the efficiencies are 85.6 and 83.5 % for

kR and *γR* procedures, respectively. For the procedure, *γN*, the fixed *γ* value for a sample of 50 families with 16 individuals and three traits, is 0.3 (HAYES & HILL 1981). The value of average expected gain corresponding to $\gamma = 0.3$ is 0.2758 giving efficiency of the procedure, *γN*, as 78.2 %. The efficiencies of other two procedures ($k = 0$ and $k = 1$) were also calculated in this way.

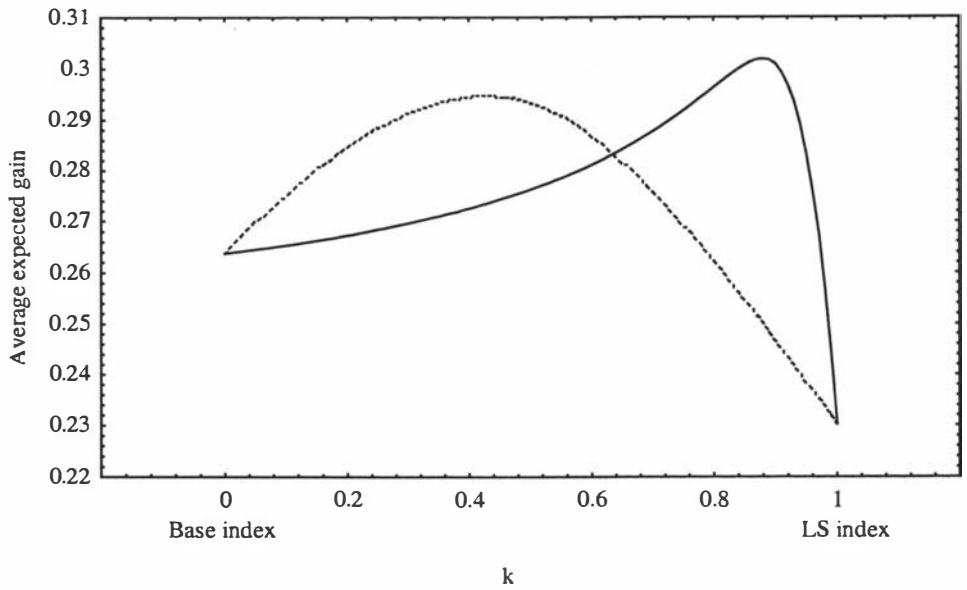


Figure 1. Average (over 1000 replications) expected gain (equation 5) for two index selection procedures *kR* (——) and *γR* (- - -), $f = 50$, $n = 15$, set 8, REV = 3,2,1. $\gamma = 1\text{-}k$.

The simulation results regarding best fixed *k* values and efficiencies of different index selection methods are given in Tables 2 to 6. The values of *k* that maximised the average expected gain for various sample situations, are given in Table 2. These values were averaged over all parameter sets given in Table 1. Lower *k* values were required for small sample size. The overall best *k* values were 0.65, 0.75, 0.84 and 0.91 for a sample of 25, 50, 100 and 200 families, respectively. This indicates that sample estimates with small number of families are less reliable. Thus, more weight should be given to REVs to overcome the effects of higher sampling bias associated with small sample size. As the sample size increases to 200 families, the estimates are less affected by sampling errors as is evident from higher *k* value (0.91) required for regressing the least-square index weights towards the REVs. The extent of sampling errors also increased with increase in number of traits included in selection index. For a sample of

25 families, the optimum k values ranged from 0.58 to 0.74 depending upon the number of traits included in index. Thus sampling errors are more important with more traits in an index. However, for a large sample size of 200 families, the number of traits had little impact on the best fixed k values.

Table 2. Average values (which maximise the expectation of genetic gain) of k for different sample sizes. Averaged over all parameter sets in Table 1.

Number of families	Number of traits			Overall
	2	3	4	
25	0.74	0.63	0.58	0.65
50	0.79	0.76	0.70	0.75
100	0.87	0.84	0.81	0.84
200	0.93	0.91	0.90	0.91
Overall	0.83	0.78	0.75	

The relative efficiencies of various index procedures for varying sample sizes are given in Table 3. The % efficiencies presented in this table are averaged over all parameter sets with zero environmental correlations. For the procedure kN, the fixed k values were taken from Table 2 and the efficiency corresponding to the fixed k values were calculated for this procedure. The relative efficiency of $k = 1$ varied from 70.9 to 95.3 % (Table 3) increasing with number of families whereas the efficiency of $k = 0$ was constant at 82.9 % as sample information contributes nothing in this procedure of index selection. The procedure $k = 0$ is better than $k = 1$ only with small sample sizes. The γR procedure was found to be about 2 to 3 % more efficient than γN across all sample sizes. The slightly lower relative efficiency of γN as compared to γR is because γR represents the maximum point on the plot of average expected gain (as shown in Figure 1). The efficiency of the kR procedure was higher than for the other procedures. The procedure kN was found to be less efficient than kR and γR but its efficiency is higher compared to γN.

Table 3. Average efficiency (%) of different index procedures for different sample sizes. Averaged over parameter sets with zero environmental correlations.

Number of families	*kR*	*γR*	*k* = 0 (Base index)	*k* = 1 (LS index)	*γN*	*kN*
25	91.2	89.8	82.9	70.9	86.6	88.6
50	93.6	93.2	82.9	82.5	90.1	91.0
100	95.6	95.5	82.9	90.5	92.9	93.5
200	97.3	97.3	82.9	95.3	95.1	96.2

The results of the effect of number of traits on the efficiency of different procedures are given in Table 4. As the number of trait increases, the efficiency of all procedures also decreases. The rate of declining efficiency is lowest for $k = 0$ whereas highest decline was observed for $k = 1$. The magnitude of the relative efficiency advantage of the procedure kR over γR declines as the number of trait increases but its superiority is maintained throughout.

Table 4. Average efficiency (%) of different index procedures for different number of traits. Averaged over parameter sets with zero environmental correlations.

Number of traits	Number of families	*kR*	*γR*	*k* = 0	*k* = 1	*γN*	*kN*
2	25	94.1	92.2	84.1	80.3	90.2	91.7
	50	96.2	95.6	84.1	88.6	93.4	93.5
3	25	90.5	88.9	83.5	67.2	85.0	88.2
	50	93.1	92.7	83.5	81.7	89.4	90.9
4	25	88.9	88.3	81.3	65.4	84.4	86.0
	50	91.6	91.2	81.3	77.3	87.4	88.6

The relative efficiency of various procedures was also evaluated for different economic weights and heritability relationships. Average relative efficiencies for this part of analysis are given in Table 5. The efficiency was higher, for all procedures, when

economic weights were in the same rank order as heritabilites. The loss in efficiency, when REVs and heritabilites are in opposite order, is least for *kR*. Also, the magnitude of higher relative efficiency of *kR* is maximum when there is opposite relationship between REVs and heritabilites while it is minimum when REVs and heritabilites are in the same order. As the procedure *kR* regressed the estimated index coefficients towards REVs, therefore, loss in efficiency is lower as compared to other procedures.

Table 5. Average efficiency (%) of different index procedures for different relationship between economic weights and heritabilites. Averaged over parameter sets with zero environmental correlations.

Economic weights : heritabilities	Number of families	*kR*	*γR*	*k* = 0	*k* = 1	*γN*	*kN*
1	25	91.4	90.1	83.5	69.2	86.2	89.0
	50	93.8	93.4	83.5	81.3	88.9	91.1
	100	95.7	95.6	83.5	89.9	91.5	93.6
	200	97.3	97.2	83.5	94.9	93.9	96.3
2	25	94.5	93.5	86.1	80.2	90.7	92.3
	50	96.3	96.1	86.1	89.6	93.6	94.5
	100	97.5	97.4	86.1	94.8	95.7	96.4
	200	98.5	98.4	86.1	97.5	97.1	98.0
3	25	87.5	85.9	79.3	63.4	82.8	84.6
	50	90.7	90.1	79.3	76.6	87.7	87.4
	100	93.7	93.5	79.3	86.7	91.7	91.2
	200	96.2	96.1	79.3	93.3	94.5	94.3

Economic weights : heritabilites = 1: Economic weights are equal for all traits
= 2: Economic weights and heritabilites are in same order
= 3: Economic weights and heritabilites are in opposite order

Prompted by the results in Table 5, it was decided to compare the efficiency of various procedures in the situation where all the index traits have low heritabilites. For this purpose parameter sets 4, 8 and 12 (see Table 1) were chosen. When economic traits

and heritabilities are in opposite rank order, the average efficiency of kR was 5.5 and 2.3 % higher than γR with a sample size of 25 and 50 families, respectively.

Table 6. Average efficiency (%) of different index procedures for different number of traits. Averaged over parameter sets with non-zero environmental correlations.

Number of traits	Number of families	kR	γR	$k = 0$	$k = 1$	γN	kN
3	25	97.2	97.1	96.3	85.9	95.9	97.0
	50	97.9	97.9	96.3	93.9	96.9	97.7
	100	98.5	98.5	96.3	97.1	98.0	98.4
4	25	96.7	96.7	95.9	80.5	95.5	96.5
	50	97.2	97.1	95.9	89.0	95.7	97.0
	100	97.9	97.9	95.9	95.1	96.8	97.8

As mentioned earlier, this study was also designed to examine the impact of non-zero environmental correlations on the relative efficiencies of various procedures. The parameter sets 13 to 16 given in Table 1 were used for this part of analysis. The results for 3 and 4 traits-index are shown in Table 6. It shows that the procedure kR has an advantage over all other procedures. One interesting result clearly apparent by comparing Table 4 and Table 6 is that the difference in the relative efficiency of procedure $k = 0$ with others has reduced dramatically in Table 6. Interestingly, $k = 0$ procedures have shown slightly higher relative efficiency as compared to $k = 1$ even with a sample size of 100 families and 4 traits-index.

DISCUSSION

The present study assumes a balanced structure of half-sib families. The effect of sampling errors on the efficiency of index selection, was evaluated using ANOVA estimates of genetic parameters. Different methods of modifying the parameter estimates to increase their reliability, have been proposed in the past. HAYES & HILL (1981) pointed out the possibility of modifying the index weights themselves which has some analogies with the technique of ridge regression (HOERL & KENNARD 1970). The proposed method is a form of index which takes into account the REVs which are

assumed to be known precisely. This method lowers the chances of reducing the efficiency of index selection when the breeder is not confident about the reliability of parameter estimates. The proposed method is also an intermediate solution between the two extreme situations, i.e. LS index and 'base index' and maximises the expected genetic gain. This method is similar, although methodologically different, to that of HAYES & HILL (1981). The expectation of response that would be actually achieved, for evaluating the efficacy of the different index selection procedures used in this study, should be theoretically similar to those from maximising the correlation between true and estimated aggregate genotype. The latter criterion has been used in some studies (e.g., ESSL 1991, VERRYN 1994).

Effect of sample size and genetic parameters

The maximum genetic gain was obtained with $k < 1$ when sample estimates of variance components were used. The optimum value of k was found to depend on number of traits, size of experiment and heritabilities. With a two-trait index, the magnitude of sampling errors is comparatively less and thus highest expected gain was obtained with k values closer to 1. With low heritabilities of index traits, the maximum gain was obtained at comparatively low k values. These findings parallel those obtained in other studies (e.g. HAYES & HILL 1981, ESSL 1991) with large γ values. The genetic parameters were found to have influence on the optimum value of bending factor in these studies. The logic behind the γN (i.e., select bending factor on the basis of sample size alone) of HAYES & HILL (1981) was to operate more generally when all roots are positive and genetic parameters are unknown. The relative sub-optimality of procedure γN compared to γR in the present study may be because of the effect of genetic parameters on optimum γ.

The calculation of expectation of response in (5) requires that the true parameters be known. However, the improvement of the efficiency of different index selection procedures can only be calculated in Monte Carlo simulation studies (MEYER & HILL 1983). By substituting $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{G}}$ in place of $\boldsymbol{P}$ and $\boldsymbol{G}$ in (5), a roughly linear decline (results not shown) in $\hat{R}$ with decreasing k was obtained. It indicates that the pattern of predicted gains gives no real guidance about optimum k, unless perhaps one is dealing

with large matrices and obviously unstable LS index solution which shows clear parallels with multiple regressions.

Effect of environmental correlations

The average expected genetic gain for kR procedure was much higher compared to the unmodified index selection (i.e. at $k = 1$). The difference in the relative efficiency of kR and $k = 1$ procedures fall drastically when non-zero environmental correlations were taken into account (Table 6). One of the probable reasons for this may be that results in Table 6 are based on parameters set where heritabilities of traits are marginally higher which resulted in the higher relative efficiency of $k = 1$. ESSL (1991) also looked briefly into the aspect of non-zero environmental correlations. Further investigation is required to study the effect of various degrees of environmental correlations on relative efficiency of different index selection procedures.

Effect of different economic weights

Precise knowledge of REVs was assumed in this study. However, the choice of appropriate economic weights can itself be crucial, particularly when adverse genetic correlations are involved. The economic end-product value of observed traits is often difficult to evaluate, especially in tree breeding programmes that involve long generation intervals and uncertain relationships between biological traits and net end-product values. The proposed method proved to be relatively more efficient, under various scenarios, as compared to other procedures considered in this study (Table 4). Its efficiency was much higher particularly when the index involves low-heritabilities traits with REVs in opposite rank order to heritabilities. The results also shows that the proposed method is more efficient than other procedures even when equal REVs have been assigned to different traits. The efficiency of γR was, however, almost identical to that of kR when sample size was more than 50 families.

In this study the efficacy of the proposed method was tested for MANOVA estimation of genetic parameters in a balanced half-sib family structure. Further investigation is required for establishing the efficacy of this method for other genetic parameter estimation methods (like REML) and selection methods (different sources of information: ancestors, individual and progeny). This study does not rule out the

possibility that better methods of modifying the parameter estimates exist. Our results offer a simple but very effective procedure which can be further explored.

ACKNOWLEDGEMENTS

S Kumar thanks N. Z. Forest Research Institute, Rotorua, New Zealand, for financial support.

REFERENCES

ARNASON, T. 1982: Prediction of breeding values for multiple traits in small non-random mating (horse) populations. *Acta Agriculturae Scandinavica* **32**:171-176.

BULMER, M. G. 1985: The Mathematical Theory of Quantitative Genetics. Clarendon Press, Oxford, 255 pp.

ESSL, A. 1991: Choice of an appropriate bending factor using prior knowledge of parameters. *Journal of Animal Breeding and Genetics* **108**:89-101.

FALCONER, D. S. 1989: Introduction to Quantitative Genetics. 3rd ed. Longman Scientific & Technical, Essex, England. 438 pp.

HARRIS, D. L. 1964: Expected and predicted progress from index selection involving estimates of population parameters. *Biometrics* **20**:46-72.

HAYES, J. F. & HILL, W. G. 1981: Modification of estimates of parameters in the construction of genetic selection indices ('Bending'). *Biometrics* **37**:483-493.

HAZEL, L. N. 1943: The genetic basis for constructing selection indices. *Genetics* **28**:476-490.

HILL, W. G. & THOMPSON, R. 1978: Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* **34**:429-439.

HOERL, A. E., & KENNARD, R. W. 1970: Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* **12**:55-67.

KEMPTHORNE, O. 1957: An Introduction to Genetic Statistics. Wiley, New York, 545 pp.

KENNEDY, W. J. &, GENTLE, J. E. 1980: Statistical Computing. Marcel Dekker, New York, 591 pp.

MEUWISSEN, T. H. E. & KANIS, E. 1988: Application of bending theory in a pig - breeding situation. *Livestock Production Science* **18**:85-91.

MEYER, K. & HILL, W. G. 1983 A note on the effect of sampling errors on the accuracy of genetic selection indices. *Journal of Animal Breeding and Genetics* **100**:27-32.

ODELL, P. L. & FEIVESON. A. H. 1966: A numerical procedure to generate a sample covariance matrix. *Journal of American Statistical Association* **61**:199-203.

SALES, J. & HILL, W. G. 1976 a: Effect of sampling errors on the efficiency of selection indices. 1. Use of information from relatives for single trait improvement. *Animal Production* **22**:1-17.

SALES, J. & HILL, W. G. 1976 b: Effect of sampling errors on the efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. *Animal Production* **23**:1-14.

SAS INSTITUTE INC 1989: SAS/IML Software: Usage and reference. Cary, NC: SAS Institute Inc, 501 pp.

SAXTON, A. M. 1986: A comparison of bending and ridge regression in selection index estimation. *In*: Dickerson, GE, Johnson, RK eds. Proc 3rd World Congress on Genetics Applied to Livestock Production, Lincon, 16-22 July 1986 , **12**:449-453.

SMITH, H. 1936: A discriminant function for plant selection. *Annals of Eugenics* **7**:240-250.

TAI, G. C. C. 1989: A proposal to improve the efficiency of index selection by 'rounding'. *Theoretical and Applied Genetics* **78**:798-800.

VERRYN, S. D. 1994: Improving on best linear prediction for tree breeding. PhD Thesis. University of Pretoria, Pretoria.

WILLIAMS, J. S. 1962 a: Some statistical properties of a genetic selection index. *Biometrika* **49**:325-337.

WILLIAMS, J. S. 1962 b: The evaluation of a selection index. *Biometrics* **18**:375-393.

XU, S. & MUIR, W. M. 1990: The application of ridge regression to multiple trait selection indices. *Journal of Animal Breeding and Genetics* **107**:81-88.

Chapter 7

GENERAL DISCUSSION

Detection and independent verification of quantitative trait loci are important steps in the process of marker assisted selection (MAS). The mapping designs greatly influence power to correctly identify QTL for a given size experiment. This chapter discusses different methods that have been put forward to calculate critical values for rejecting or failing to reject the alternative hypothesis regarding the existence of a QTL; and reviews different options for establishing the experimental designs for QTL mapping. In addition, the implementation of MAS in radiata pine breeding is discussed.

Full-sib families are commonly being used as QTL mapping populations in forest trees. Being outbred populations, forest trees are highly heterozygous and there might be more than two alleles segregating at any locus. Therefore, the methods developed for establishing QTL mapping designs in populations derived from inbred lines can not be used for considering the designs for QTL detection in forest trees. Statistical methods used to detect QTL generally suppose that QTL act on the phenotypic trait mean. In the simpler approach (e.g. SOLLER *et al.*, 1976), marker genotypes are considered as factors in a one-way ANOVA, and the test for significance of this factor is the test for the presence of a QTL in the vicinity of the marker locus. The theory for single-marker ANOVA, when male and female parents are segregating different marker alleles, was developed (Chapter 2) to help determine the size of experimental designs for detecting QTL.

In Chapter 2, the results showed that the power of linkage detection was higher for fully-informative (FI) marker designs compared to pseudo-backcross (PBC) and pseudo-intercross (PIC) marker configurations. This result can be explained in terms of the expected number of informative marker contrasts and the number of informative offspring in each case. For a given QTL size, the methodology presented in Chapter 2 and 3 can be used to determine the optimum number of families to achieve maximum power of QTL detection. However, the question that how many families (independent or in some mating design structure) are required to find all QTL of large effects in the population need to be answered.

The methodology presented in Chapter 2 and Chapter 3 provides a useful tool to enable quick screening of different scenarios (experimental and genetical) deterministically before establishing a QTL mapping trial in outbred populations. LANDER & BOTSTEIN

(1989) suggested that power of detecting marker-QTL linkages could be markedly increased by utilising interval mapping with the likelihood ratio test as compared to single-marker ANOVA tests. KNOTT *et al.* (1996), using stochastic simulations for outbred half-sib pedigrees, showed that the single-marker method provides the lower limit of the power and it can be increased using multiple-marker methods. FU & RITLAND (1994) studied the statistical properties of mapping recessive viability loci using the single-marker approach. They showed that two flanking markers provide vastly superior estimation properties and required smaller sample sizes compared to those required for equivalent power utilising a single marker.

DARVASI *et al.* (1993), similar to HALEY & KNOTT (1992), found that the difference in power between interval mapping using a likelihood ratio test and single-marker ANOVA was small. When intervals of up to 20 cM are used, they found little difference in the results obtained using two methods. They argue that the comparison made by LANDER & BOTSTEIN (1989) did not take into consideration that when a pair of flanking markers is available, both will be individually examined in the corresponding single-marker analysis. Statistical significance with respect to either will result in marker-QTL linkage identification, hence increasing the power of single-marker analysis. Also, LANDER & BOTSTEIN (1989) only investigated the case where the QTL is located at the mid-point with respect to the flanking markers. This is the worst case for single-marker QTL linkage determination relative to interval mapping. In practice, single-marker analysis can be used for preliminary screening for any marker-trait association and then multi-marker methods should be used for determining the position and effect of QTL.

The disadvantage of the single-marker ANOVA approach is that it gives no information on the position of the QTL as it cannot separate its quantitative effects from its distance from the marker (MURANTY, 1996). Methods using flanking markers (LANDER & BOTSTEIN, 1989; HALEY & KNOTT, 1992) allow separate estimation of QTL position and effect. HALEY *et al.*, 1994 pointed out that interval mapping may be biased towards locating a QTL in the most informative interval rather than the correct one and thus, suggested the simultaneous use of all markers in a linkage group (popularly known as multiple marker mapping). HALEY *et al.* (1994) and KNOTT *et al.* (1996) extended the use of multiple marker mapping to outbred populations. This approach was used in

Chapter 4 for analysing the marker-trait information from linkage group three of radiata pine.

Threshold levels

A number of statistical methods are being used for detecting and locating QTL. A common problem associated with all of these methods is the determination of appropriate significance thresholds (critical values) against which to compare test statistics for determining the presence of a QTL. In Chapter 2 and 3, the power of linkage detection was obtained at a type-I error rate of 0.01. In practice, a number of QTLs may be missed (a type-II error) and at the same time a number of false positives may occur (a type-I error). Keeping type-I error rate at, say, 0.05 the power could have increased. The actual balance between the cost of false positives and the benefit of detected QTL depends on the aim of the experiment. If the objective of the experiment is to identify QTL that will be subsequently confirmed in an independent population, a type-I error rate of 0.15-0.20 may be used to ensure QTL are not missed.

The determination of what appropriate significance thresholds should be used is very critical. As discussed by CHURCHILL & DOERGE (1994), there are two sources of this problem. First, there is the problem of determining (or approximating) the distribution of the test statistic under an appropriate null hypothesis. The distribution of the null hypothesis has been approximated through theoretical methods (LANDER & BOTSTEIN, 1989; FEINGOLD *et al.*, 1993; REBAI *et al.*, 1994), as well as empirical methods (CHURCHILL & DEORGE, 1994). The reliability of the asymptotic (theoretical) approximation is doubtful because of problems such as finite sample sizes and distributional properties of the trait and thus, empirical approaches should be used (CHURCHILL & DEORGE, 1994). The empirical method of CHURCHILL & DOERGE (1994), applied in Chapter 4, also inherently accounts for other characteristics of the data set (e.g. missing phenotypic or genotypic data, segregation distortion), whereas the theoretical approximations are based upon "perfect" data.

The second source of difficulty is the multiple hypothesis testing that is implicit in genome searches used for locating QTL. A large number of tests (e.g. every cM) may be carried out and a number of traits may be considered, many of which are not independent. As reviewed by SPELMAN (1998), three methods that address the effect of

multiple testing have been put forward for calculating critical values. These methods are: experimental type I error rate (LANDER & BOTSTEIN, 1989; LANDER & KRUGLYAK, 1995), false discovery rate (BENJAMINI & HOCHBERG, 1995; WELLER *et al.*, 1998), and posterior type I error (SOUTHEY & FERNANDO, 1998). All these methods to set significance levels have some drawbacks, but the method of setting experimentwise levels (LANDER & BOTSTEIN, 1989; LANDER & KRUGLYAK, 1995) appears to be most applicable to genome scans with interval mapping (SPELMAN, 1998).

Design of QTL mapping experiments

Type and structure of mapping population

A QTL mapping experiment should be carefully designed. Full-sib families are being widely used in QTL mapping studies in forest trees (e.g. GROOVER *et al.*, 1994; KNOTT *et al.*, 1997; EMIBIRI *et al.*, 1998; KUMAR *et al.*, 2000). Chapter 2 of this thesis considers the power of QTL detection experiments using full-sib families. Use of full-sib families in conifers is directed at detecting QTL segregating in the normal course of outbreeding within a population. PLOMION *et al.* (1996) used a self-family for detecting QTL for early growth. Detection of QTL in self-families would not be the same as in outbred pedigrees. Some of the QTL in self-families will represent genetic load (resulting in loss of both survival and reproduction); however, there would be some individuals without appreciable load so QTL detected in these self pedigrees would be similar to those expressed under the normal course of outbreeding (PLOMION *et al.*, 1996). In Chapter 3 of this thesis, the methodology for estimating the power of QTL detection in self-families of outbred populations is presented.

Many factors, including size and structure of the mapping population, influence the efficiency of QTL mapping experiments. As shown in Chapter 2 and Chapter 3 and by many other authors (e.g. HILL, 1975; SOLLER & GENIZI, 1978; WELLER *et al.*, 1990; VAN DER BEEK *et al.*, 1995), increasing the number of offspring per family is more efficient with respect to power than increasing the number of families. A question of importance is whether the tree breeding population as it exists has a size and structure suitable for QTL analysis. The size of full-sib families in tree breeding programs is not large enough (<100) for powerful detection of small to moderate size QTL. However, there might be some large families available in commercial plantations that can be used

for this purpose but the use of these families introduces the risk of mislabelling of offspring.

Selective genotyping

Power of QTL detection shown in Chapter 2 and Chapter 3 are based on the assumption that all the available offspring in any pedigree are genotyped. The results showed that large sample sizes are required for detecting small to moderate size QTL. Many traits are less expensive to score than marker genotypes. In such cases, if our interest is a single trait (as opposed to experiments where a large number of traits are simultaneously considered), it pays to first score a number of individuals for the trait and then genotype only a selected subset of these. This strategy is called selective genotyping and can result in a large increase in power for a given number of genotypes (LEBOWITZ *et al.*, 1987; DARVASI & SOLLER, 1992). The basis of this approach is that much of the linkage information resides in individuals with extreme phenotypes.

As shown in Table 1 and Table 2, family size has an impact on the power of detection, as the larger the family size, the greater the power. Complete genotyping certainly gives higher power compared to selective genotyping. However, most of the power of linkage detection can be achieved by selectively genotyping 5-10 % of the total offspring from each tail of the trait distribution. The decision about the optimum proportion to be genotyped from each tail is also influenced by the trait heritability (Table 1 and Table 2). For low heritability traits, a higher proportion would have to be genotyped to get reasonably close to the maximum achievable power from an experiment. A smaller proportion selected from the tails of the phenotypic distribution of low heritability traits might be quite misleading because of the higher environmental contribution towards the phenotype. The best strategy for future QTL detection experiments would be to get larger family sizes and selectively genotype offspring from the tails of the phenotypic distribution.

Asexual propagation

As asexual propagation is not difficult in radiata pine, it offers another QTL mapping design strategy for increasing power (LANDER & BOTSTEIN, 1989; SOLLER & BECKMANN, 1990; KNAPP & BRIDGES, 1990). The idea is to reduce the effects of

Table 1 Power of QTL detection using selective genotyping for a single full-sib family. The QTL explains 5 % of the genetic variance (or 0.27 σ_p, this is '*a*' in Falconer terms) for a trait with heritability 0.75. Recombination between marker and the QTL is assumed to be 0.10.

# selectively genotyped	Family size			
	500	1000	2000	4000
100	0.32	0.42	0.49	0.56
200	0.42	0.59	0.64	0.70
300	0.48	0.66	0.69	0.74
All	0.48	0.70	0.73	0.75

Table 2 Power of QTL detection using selective genotyping for a single full-sib family. The QTL explains 5 % of the genetic variance (or 0.16 σ_p, this is '*a*' in Falconer terms) for a trait with heritability 0.25. Recombination between marker and the QTL is assumed to be 0.10.

# selectively genotyped	Family size			
	500	1000	2000	4000
100	0.06	0.10	0.12	0.14
200	0.08	0.15	0.20	0.25
300	0.09	0.17	0.24	0.38
All	0.09	0.23	0.47	0.72

environmental variation by asexually replicating each genotyped individual, using the mean value of these replicated progeny in place of the single individual value. This is not an efficient strategy unless the trait heritability is small, in which case scoring only a few replicated progeny can result in a significant increase in power. SOLLER & BECKMANN (1990) showed that most of the increase in the power occurs by measuring ten or fewer replicated progeny. Besides offering some increase in power, progeny replication allows marker-trait associations to be examined across environments, allowing QTL-environment interactions to be estimated (LYNCH & WALSH, 1997).

Marker Assisted Selection (MAS)

An outbred population is a collection of families and the linkage disequilibrium between markers and QTLs differs over families, i.e., the outbred population is expected to be in linkage equilibrium. Thus, marker allelic effects have to be estimated for each family separately. Genetic gains predicted from MAS in crossbred populations are quite high (e.g. LANDE & THOMPSON, 1990). However, the estimated marker allelic effects in an outbred population explain less variance and therefore contribute less to genetic gain. A single family in an outbred population is usually smaller than an entire crossbred population. Therefore, the marker allelic effect estimates are regressed more towards zero in an outbred population (VAN DER BEEK, 1996). It is less likely to find large segregating QTLs in an outbred population compared to crossbred population.

In other plants and animal species, many authors have looked at the implications of MAS on breeding programmes through simulation. However, there are not many published reports on the application of MAS in forest tree breeding programmes. MAS can be applied in two different ways in tree breeding programmes. First, within-family MAS, where selection decisions are first made on estimated breeding values and then within-family selection decisions are based on QTL information. Second, across-family MAS, where selection decisions are made on breeding values that combine the QTL and the polygenic components (FERNANDO & GROSSMAN, 1989). However, across-family MAS does not seems to be a practical option in the near future.

KERR *et al.* (1996) simulated a multi-generation across-family MAS scheme and reported that after 3 generations of selection and breeding the relative genetic gains in total aggregate genetic merit using MAS would be 107-116 % for breeding population and 106-113 % for deployment population. Some other unpublished studies (G. R. Johnson *et al.*) simulated MAS and the genetic gains in deployment populations were shown to be promising. The results presented in Chapter 5 of this thesis shows that genetic gain in conventional selection and deployment schemes can be increased using MAS. Genetic gain improvement achieved with MAS is dependent on the genetic model considered. In reality, the underlying genetic model is not known i.e. the number of alleles, distribution of effects, and interaction between loci. Because of this, different types of genetic models are being simulated in MAS studies: varying number of QTLs, sizes of the QTL, ranging from bi-allelic QTL to many alleles at the QTL. In any case,

the extra gain achieved through MAS needs to be evaluated in economic terms. A formal economic evaluation of MAS would need to take into account the extra costs associated with marker typing and the time taken for extra evaluation. Assuming different genetic models, some unpublished studies (G. R. Johnson *et al.*; P. L. Wilcox *et al.*) have shown economic benefits of MAS under various deployment options.

Implementation of MAS in radiata pine breeding in New Zealand

Marker-QTL information can be found for all parents that are being used in production population. This marker-trait information can be used to increase genetic gain by selecting within-families that are used to establish production plantations, and propagating for deployment only those individuals that have favourable marker haplotypes. Elite populations are also being established for some economically important traits. All elite families can be screened for marker-trait associations and MAS can be used in these elite families to select individuals for the next generation breeding and also to become the parents of the production populations. In radiata pine breeding in New Zealand, clonal selection is being practised to capture the non-additive genetic component particularly for low heritability traits. Generally, clonal selection is carried out in full-sib families of genetically superior parents (determined by their progeny test). This is also the potential area in tree breeding where MAS can be applied for selecting only those genotypes, at seedling stage, which have the desired haplotype for putting them in clonal test. After field testing of MAS clones, superior clones can be selected for deployment in the production plantations.

Index Selection

Across-family MAS is technically more demanding than within-family MAS, but genetically superior to within-family MAS. This is because the inclusion of QTL information in the estimation of breeding values results in more accurate estimation and therefore higher selection differentials as differences between families can be exploited as well as within (SPELMAN & GARRICK, 1998). KERR & GODDARD (1997) suggested that conventional options should be first optimised for increasing genetic gain. These conventional options might include numbers of families, family size and some other measures to reduce the effect of sampling errors on the estimation of genetic parameters. SALES & HILL (1976) and HAYES & HILL (1981) investigated the effect of genetic parameter estimation error on selection response and suggested some measures

to improve the efficiency of selection and thus increasing the genetic gain in breeding programmes. In Chapter 6 of this thesis, a method is proposed to reduce the effect of sampling errors on the selection index coefficients and thus increase the efficiency of index selection. The procedure, *kN*, given in Chapter 6 can be used in practice for scaling the estimated index weights towards economic weights. The choice of scaling parameter, *k*, will depend on the sample size and the number of traits being considered in the selection index. Most of the MAS studies have dealt with a single trait only. However, in practice, the commercial breeding objective is always composed of several traits. WELLER *et al.* (1988) found QTL with positive effects on pairs of traits. Selection for these loci may be significantly more effective than the traditional selection index (e.g. DE KONING & WELLER, 1994). The application of the method given in Chapter 6 was explored only for the traditional selection index and thus its efficiency in case of multitrait MAS remains to be evaluated.

REFERENCES

BENJAMINI, Y. & HOCHBERG, Y. 1995: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society* **57(1)**:289-300.

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.

DARVASI, A. & SOLLER, M. 1992: Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics* **85**:353-359.

DE KONING, G. J. & WELLER, J. I. 1994: Efficiency of direct selection on quantitative trait loci for a two-trait breeding objective. *Theoretical and Applied Genetics* **88**:669-677.

EMEBIRI, L. C., DEVEY, M. E., MATHESON, A. C. & SLEE, M. U. 1998: Age-related changes in the expression of QTLs for growth in radiata pine seedlings. *Theoretical and Applied Genetics* **97**:1053-1061.

FEINGOLD, E., BROWN, P. O. & SIEGMUND, D. 1993: Gaussian models for genetic linkage mapping using complete high-resolution maps of identity by descent. *American Journal of Human Genetics* **53**:234-251.

FERNANDO, R. L. & GROSSMAN, M. 1989: Marker-assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* **21**:467-477.

FU, Y. B. & RITLAND, K. 1994: On estimation of linkage of marker genes to viability genes controlling inbreeding depression. *Theoretical and Applied Genetics* **88**: 925-932.

GROOVER, A., DEVEY, M., FIDDLER, T., LEE, J., MEGRAW, R., MITCHEL-OLDS, T., SHERMAN, B., VUJCIC, S., WILLIAMS, C. & NEALE, D. 1994: Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**:1293-1300.

HILL, A. P. 1975: Quantitative linkage: a statistical procedure for its detection and estimation. *Annals of Human Genetics* **38**:439-449.

HALEY, C. S. & KNOTT, S. A. 1992: A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**:315-324.

HALEY, C. S., KNOTT, S. A. & ELSEN, J. M. 1994: Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**:1195-1207.

HAYES, J. F. & HILL, W. G. 1981: Modification of estimates of parameters in the construction of genetic selection indices ('Bending'). *Biometrics* **37**:483-493.

KERR, R. J., JARVIS, S. R. & GODDARD, M. E. 1996: The use of genetic markers in tree breeding programs. Pp. 498-505 *in*: M. J. DIETERS, A. C. MATHESON, D. G. NICKLES, HARWOOD, C. E. & MALKER, S. M. (eds). *Tree Improvement for Sustainable Tropical Forestry*. Proc. QFRI-IUFRO Conference, Caloundra, Queensland, Australia, 27 Oct – 1 Nov 1996.

KERR, R. J. & GODDARD, M. E. 1997: A comparison between the use of MAS and clonal test in breeding programs. Pp. 297-303 *in*: BURDON, R. D. & MOORE, J. M. (eds). *IUFRO '97 Genetics of Radiata Pine*. Proc. NZ FRI-IUFRO Conference 1-4 Dec and Workshop Dec 5, Rotorua, New Zealand. FRI Bulletin No. 203.

KNAPP, S. J. & BRIDGES, W. C. 1990: Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* **126**:769-777.

KNOTT, S. A., ELSEN, J. M. & HALEY, C. S. 1996: Methods for multiple marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**: 71-80.

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.

KUMAR, S., SPELMAN, R. J., GARRICK, D. J., RICHARDSON, T. E., LAUSBERG, M. & WILCOX, P. L. 2000: Multiple marker mapping of wood density loci in an outbred pedigree of radiata pine. *Theoretical and Applied Genetics* **100**:926-933.

LEBOWITZ, R. J., SOLLER, M. & BECKMANN, J. S. 1987: Trait-based analysis for the detection of linkage between marker locus and quantitative trait loci in crosses between inbred lines. *Theoretical and Applied Genetics* **73**:556-562.

LANDE, R. & THOMPSON, R. 1990: Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**:743-756.

LANDER, E. S. & BOTSTEIN, D. 1989: Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**:185-199.

LANDER, E. & KRUGLYAK, L. 1995: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**:241-247.

LYNCH, M. & WALSH, B. 1997: Genetics and Analysis of Quantitative Traits. Sinacer Associates, Massachusetts, USA. 980 p.

MURANTY, H. 1996: Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**:156-165.

PLOMION, C., DUREL, C.-E. & O'MALLEY, D. M. 1996: Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. *Theoretical and Applied Genetics* **93**:849-858.

REBAI, A., GOFFINET, B. & MANGIN, B. 1994: Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**:235-240.

SALES, J. & HILL, W. G. 1976: Effect of sampling errors on the efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. *Animal Production* **23**:1-14.

SOLLER, M., GENIZI, A. & BRODY, T. 1976: On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical and Applied Genetics* **47**:35-39.

SOLLER, M. & GENIZI, A. 1978: The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* **34**:47-55.

SOLLER, M. & BECKMANN, J. S. 1990: Marker based mapping of quantitative trait loci using replicated progenies. *Theoretical and Applied Genetics* **80**:205-208.

SOUTHEY, B. R. & FERNANDO, R. L. 1998: Controlling the proportion of false positives among significant results in QTL detection. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia* **26**:221-224.

SPELMAN, R. J. 1998: Detection and utilisation of quantitative trait loci in dairy cattle. *PhD Thesis*, Wageningen Agricultural University, Wageningen, The Netherlands.

SPELMAN, R. J. & GARRICK, D. J. 1998: Genetic and economic responses for within-family marker assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science* **81**:2942-2950.

VAN DER BEEK, S. 1996: The use of genetic markers in poultry breeding. *PhD Thesis*, Wageningen Agricultural University, Wageningen, The Netherlands.

VAN DER BEEK, S., VAN ARENDONK, J. A. M. & GROEN, A. F. 1995: Power of two-and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theoretical and Applied Genetics* **91**:1115-1124.

WELLER, J. I., SOLLER, M. & BRODY, T. 1988: Linkage analysis of quantitative traits in an interspecific cross of tomato by means of genetic markers. *Genetics* **118**:329-339.

WELLER, J. I., KASHI, Y. & SOLLER, M. 1990: Power of daughter and granddaughterdesigns for determining linkage between marker locus and quantitative trait loci in dairy cattle. *Journal of Dairy Science* **73**:2525-2537.

WELLER, J. I., SONG, J. Z., HEYEN, D. W., LEWIN, H. A. & RON, M. 1998: A new approach to the problem of multiple comparisons for detection of quantitative trait loci. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia* **26**:229-232.

SUMMARY

This thesis focuses on the detection of quantitative trait loci (QTL) and the use of genetic markers linked to QTL in tree breeding schemes. QTL mapping experiments need to be optimised to minimise costs of data collection and genotyping. In Chapter 2, the expressions for variance components for single-marker ANOVA were derived to help determine the size of experimental designs for detecting QTL in full-sib families where parents are assumed to heterozygous for different marker alleles. The effects of experimental and genetical parameters on the power of linkage detection were evaluated. For a given experimental size, fewer larger families were better than many small families. Recombination rate, size and dominance properties of the QTL were found to have large effects on the power of linkage detection.

Molecular markers are being increasingly used in self-families of outbred organisms to find genes causing inbreeding depression in growth traits. Some of the QTL in self-families will represent genetic load; however, there would be some individuals without appreciable load so QTL detected in these self-families would be similar to those expressed under the normal course of outbreeding. In Chapter 3, the methodology for estimating the power of QTL detection in self-families of outbred populations is presented. The power of linkage detection was calculated for a wide range of progeny sizes and genetic parameters at the QTL. The magnitude of the gene effect and the recombination rate were found to have enormous effect of the power of linkage detection.

An experiment involving a full-sib pedigree (850.055 x 850.096) was established at *Forest Research* to determine the genetic location and effects of genomic regions controlling wood density at three stages (rings corresponding to ages 1-5, 6-10 and outer wood density). In Chapter 3, analysis of only a single linkage group (*three*) of the parent 850.055 was considered for mapping QTL. A multiple marker least-square approach was employed for mapping QTL for each of the three traits. Experimentwise critical values, which accounted for the evaluation of marker-QTL associations across the whole genome and for three correlated traits, were calculated. Logistic regression was used for multiple-trait QTL mapping. A putative QTL with large effect on juvenile wood density (rings corresponding to ages 1-5) appears to be segregating at 73 cM position (experimentwise $P<0.01$). The width of the 95 % bootstrap confidence interval for the putative QTL was 40 cM (i.e. 56-96 cM).

In Chapter 5, stochastic simulation was undertaken to investigate the genetic response from within-family MAS. Genetic gains obtained using MAS were compared with those obtained from conventional strategies for three selection and deployment options: 'full-sib family forestry', 'clonal forestry' and 'forwards selection for deployment'. The genetic model contained polygenes and a QTL linked to a marker. Heritability of the trait was assumed to be either 0.25 or 0.75. A QTL that explained 20 % of the genetic variance was used. It was assumed that linkage phase and the haplotype in the base population was known in order to simulate transmission of haplotype. Relative genetic gain due to the use of marker was 6-8 % and 2-3 % for 'full-sib family forestry' and 'clonal forestry' options, respectively.

Index selection is a tool commonly employed in tree breeding programs. In Chapter 6, a method is proposed to reduce the effect of sampling errors on the estimates of multivariate genetic parameters, thus increasing the efficiency of index selection. The proposed method consists of 'regressing' the estimated (least-squares) selection index coefficients towards the relative economic values. Using Monte Carlo simulations, the efficiency of the proposed method was found to be very high when the number of families used in the progeny tests are few. When the heritability of index traits was low and their relative economic values were in opposite order to heritability, the efficiency of the proposed method was much higher. The application of the proposed method was explored only for the traditional index selection and its efficiency for multitrait MAS remains to be evaluated.

In the general discussion, different QTL analysis methods i.e., single-marker ANOVA, interval mapping and multiple marker methods are outlined and discussed. Calculating the threshold levels in QTL mapping studies is always crucial. Different approaches to calculate critical values are outlined. Various aspects of designs of QTL mapping experiments in forest trees are discussed. Different strategies for increasing the power of marker-QTL linkage detection are outlined and discussed. It was shown that most of the power of linkage detection could be achieved by selectively genotyping 5-10 % of the total offspring from each tail of the trait distribution. Implementation of MAS in radiata pine breeding program in New Zealand is also discussed. Different selection and deployment scenarios are outlined where MAS can be applied in the near future.

Curriculum vitae

Satish Kumar was born on May 21st 1965 in Bakheta, Rohtak, India. He completed secondary school education in 1981 from S. M. Hindu High School, Sonepat, India. He completed his tertiary education, B. Sc. (Honours) Agriculture, in 1986 from Haryana Agricultural University, Hisar, India. During 1986 to 1988, he studied at G. B. Pant University of Agriculture and Technology, Pantnagar, India, and obtained M. Sc. (Agricultural Statistics) while receiving a Graduate Research Assistantship. From 1989 to 1994, Satish was employed as Assistant Statistician at University of Horticulture and Forestry, Solan, India. In October 1994, he was awarded a scholarship by International Tropical Timber Organisation to travel to New Zealand Forest Research Institute, Rotorua. In 1995, Satish was employed as Statistical Geneticist at New Zealand Forest Research Institute. In March 1996, he was awarded a Doctoral Scholarship by New Zealand Forest Research Institute to undertake his PhD at Massey University, New Zealand. He has been awarded a scholarship jointly by French Ministry of External Affairs and New Zealand Forest Research Institute to undertake a three-month post-doctoral period at the INRA Station de Recherches Forestieres de Bordeaux, Pierroton, France.