Acknowledgements

This research was supported in part by a grant from Korea Science and Engineering Foundation (96-0500-006-2) to MGC, a post-doc. award from Korea Science Foundation from 1996 to 1997 to JMC, and a McIntire-Stennis grant (#1774) to BKE.

References

Berg, E. E. and Hamrick, J. L.: Spatial and genetic structure of two sandhills oaks: *Quercus laevis* and *Quercus margaretta* (Fagaceae). Amer. J. Bot. **81**, 7–14 (1994). — Cheliak, W. M. and Pitel, J. P.: Technique for starch gel electrophoresis of enzyme from forest tree species. Petawawa National Forestry Institute, Information Report PI-X-42 (1984). — Chung, M. G. and Park, K. B.: Spatial genetic structure in populations of *Hosta capitata* and *H. minor* (Liliaceae). Isr. J. Plant Sci. **46**, 181–187 (1998). — Cliff, A. D. and Ord, J. K.: Spatial processes-methods and applications. Pion, London (1981). — Epperson, B. K.: Recent advances in correlation analysis of spatial patterns of genetic variation. Evol. Biol. **27**, 95–155 (1993). — Epperson, B. K., Huang, Z. and Li, T.: Measures of spatial structure in samples of genotypes for multiallelic loci. Genet. Res. Cambridge **73**, 251–261 (1999). — Hamrick, J. L., Godt, M. J. W. and Sherman-Broyles, S. L.: Factors influencing levels of genetic diversity in woody plant species. New Forests **5**, 95–124 (1992). — Haufler, C. H.: Enzyme variability and modes of evolution in *Bommeria* (Pteridaceae). Syst. Bot. **10**, 92–104 (1985). — Kamitani, T., Kawaguchi, Y., Miguchi, H. and Takeda, H.: The role of *Rhus trichocarpa* Miq. patches in pine forest dynamics in Japan. Jour. Sustainable Forestry **6**, 175–186 (1998). — Levin, D. A. and Kerster, H. W.: Gene flow in seed plants. Evol. Biol. **7**, 139–220 (1974). — Maki, M. and Yahara, T.: Spatial structure of genetic variation in a population of the endangered plant *Cerastium fischerianum* var. *molle* (Caryophyllaceae). Genes Genet. Syst. **72**, 239–242 (1997). — Matsuoka, S. and Kojima, K.: Winter food habits of grey-headed green wood peckers. Tori **28**, 107–116 (1979). — Mitton, J. B., Linhart, Y. B., Sturgeon, K. B. and Hamrick, J. L.: Allozyme polymorphisms detected in mature needle tissue of ponderosa pine. J. Hered. **70**, 86–89 (1979). — Nei, M.: Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. USA. **70**, 3321–3323 (1973). — Nei, M.: *F*-statistics and analysis of gene diversity in subdivided populations. Ann. Human Genet. **41**, 225–233 (1977). — Perry, D. J. and Knowles, P.: Spatial genetic structure within three sugar maple (*Acer saccharum* Marsh.) stands. Heredity **66**, 137–142 (1991). — Sakai, A. K. and Oden, N. L.: Spatial pattern of sex expression in silver maple (*Acer saccharium* L.): Morista's index and spatial autocorrelation. Am. Nat. **122**, 489–508 (1983). — Sherman-Broyles, S. L., Gibson, J. P., Hamrick, J. L., Bucher, M. A. and Gibson, M. J.: Comparisons of allozyme diversity among rare and widespread *Rhus* species. Syst. Bot. **17**, 551–559 (1992). — Sokal, R. R.: Ecological parameters inferred from spatial correlograms. In: Contemporary quantitative ecology and related ecometrics. Edited by G. P. Petil and M. L. Rosenzweig. International Cooperative Publishing House, Fairland. pp. 167–196 (1979). — Sokal, R. R. and Oden, N. L.: Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10**, 199–249 (1978). — Sokal, R. R. and Rohlf, F. J.: Biometry. Second edition. W. H. Freeman and Co., San Francisco (1981). — Sokal, R. R. and Wartenberg, D. E.: A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics **105**, 219–237 (1983). — Soltis, D. E., Haufler, C. H., Darrow, D. C. and Gastony, G. J.: Starch gel electrophoresis of ferns: A compilation of grinding buffers, gel and electrode buffers, and staining schedules. Amer. Fern J. **73**, 9–27 (1983). — Weeden, N. F. and Wendel, J. F.: Genetics of plant isozymes. In: Isozymes in plant biology. Edited by D. E. Soltis and P. S. Soltis. Dioscorides Press, Portland. pp. 46–72 (1989). — Workman, P. L. and Niswander, J. D.: Population studies on southern Indian tribes. II. Local genetic differentiation in the Papago. Amer. J. Hum. Genet. **22**, 24–49 (1970). — Wright, S.: The genetic structure of populations. Ann. Eugen. **15**, 313–354 (1951).

Marker-QTL Linkage Detection in Self-Families of Outbred Populations

By S. Kumar[1)2)], R. D. Burdon[1)] and D. J. Garrick[2)]

(Received 2nd April 1999)

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 and 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.

[1)] Forest Research, Private Bag 3020, Rotorua, New Zealand.

[2)] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand.

Corresponding author: Satish Kumar, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

FAX :+64 6350 5699
Tel.: +64 6350 6178
E-mail: *S. Kumar@massey.ac.nz.*

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 and SAVOLAINEN, 1996). Many studies have reported that selfed seedlings have higher mortality than outcrossed seedlings (YADZANI et al., 1985; PLESSAS and 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 and CHARLESWORTH, 1987; HUSBAND and SCHEMSKE, 1996). WILCOX (1983) reported no ID for wood density and bole straightness in selfed offspring of some parents (measured at 7 to 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 and LIBBY, 1974; LINDGREN, 1975; WRIGHT, 1980; WILCOX, 1983; WILLIAMS and 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 and 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 and 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., 1998). 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 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 and MUONA, 1990; Fu and 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 and MUONA (1990) used a single-marker approach to detect and characterise viability alleles in Scots pine (*Pinus sylvestris* L.). FU and 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 reducing 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 and 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 and 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^2_α, σ^2_β and σ^2_e, 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

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 represents 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

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.* Var (i, j) represents variance within marker genotype M_iM_j. The p and q represents 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 variance 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 *table 3.* E_{ij} represents the expected value of offspring having marker genotype M_iM_j. The p and q represents 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	$\frac{1}{2}(1-2r)^2a^2 + \frac{1}{4}(1-2r)^4d^2$
Q2Q1	½ d	$\frac{1}{2}(1-2r)^2a^2 + \frac{1}{4}(1-2r)^4d^2$
Q2Q2	-a	0

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)*. 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)*.

Using the results from *table 3, 4* and *5,* the expected variances: between families (σ^2_α), between marker genotypes within-families (σ^2_β), within-marker genotype within-families (σ^2_e), and total phenotypic variance (σ^2_T) were derived as:

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

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

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

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

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^2_β) 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_0), 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-variable. *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.

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_0) 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 a 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).

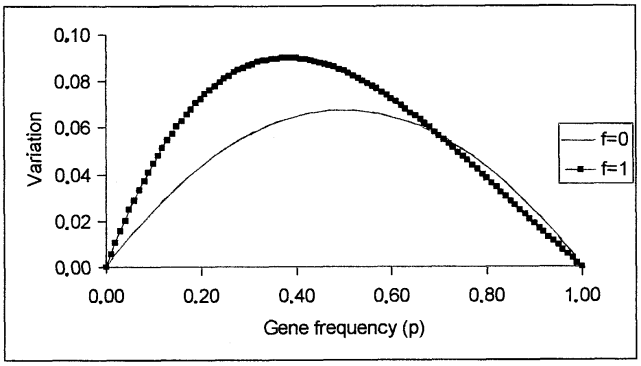

Figure 1. – Change in the contribution made by a gene to the variation of a breeding population according to the frequency and gene action. Half the difference between QTL homozygotes (i.e., a) = 0.30 SD. The dominance ratio (f) = 0 and 1 represents additive and dominant gene action, respectively.

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. 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 and 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 6. – Expected variance ratios in the single self-family with the probabilites 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

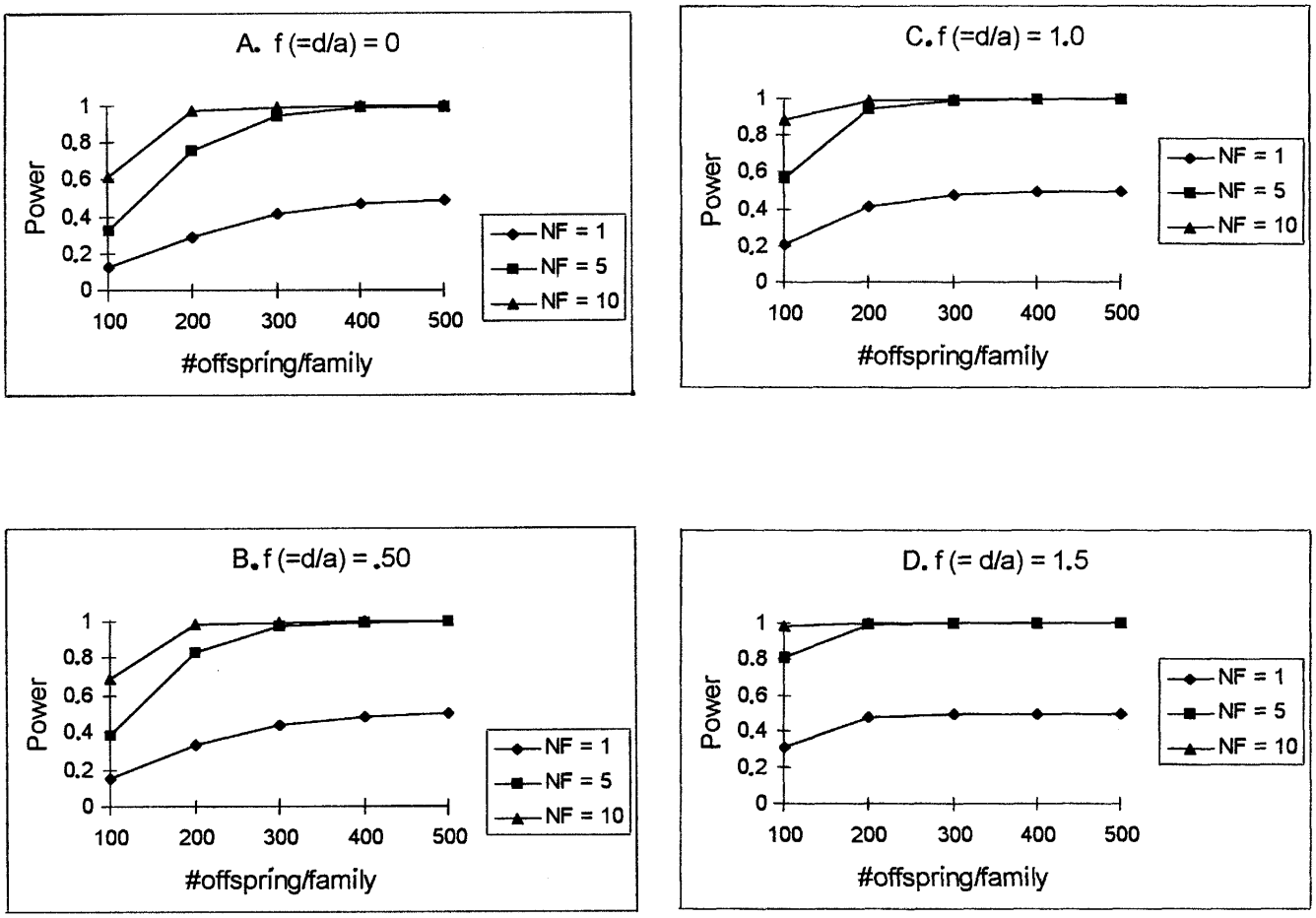

Figure 2. – The power of marker-QTL linkage detection. Different gene actions considered are: A: Additive, B: Partial-dominance, C: Dominant, D: Over-dominance. The other assumptions are: p (allele frequency of dominant allele, Q_1) = 0.5, recombination rate (r) = 0.0, Type 1 error rate = 0.01, half the difference between QTL homozygotes (i.e., a) = 0.30 SD and NF = number of self-families.

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 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 as compared to when the QTL is about 10 cM (r = .10) away from the marker. A dense

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

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

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. 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. Molecular marker linked with 'non lethals' ID-affecting locus (performance locus) should not show segregation distortion unless it is also linked to 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 and RITLAND (1994b) found performance genes contributing to the ID in vigour traits in self-families of *M. gutatus*. 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 the 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 and LIBBY, 1974; LINDGREN, 1975; WRIGHT, 1980; WILLIAMS and 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 and 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 and RITLAND, 1994b). Using these marker-trait associations, self-families can be culled for inferior individuals, at a very early stage (depend-

ing 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 and 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 and 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 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 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.

Literature

BARKER, J. E. and LIBBY, W. J.: The use of selfing in selection of forest trees. Journal of Genetics **61**: 152–168 (1974). — BURDON, R. D. and RUSSELL, J. H.: Inbreeding depression in selfing experiments: Statistical issues. Forest Genetics **5**(3): 179–189 (1998). — CHARLESWORTH, D. and CHARLESWORTH, B.: Inbreeding depression and its evolutionary consequences. Annual Review of Ecological Systems **18**: 237–268 (1987). — FRANKLIN, E. C.: Survey of mutant forms. USDA Forest Service Research Paper SE-61. 21 p. (1970). — FRANKLIN, E. C.: Genetic load in loblolly pine. American Naturalist **106**: 262–265 (1972). — FU, Y. B. and RITLAND, K.: On estimation of linkage of marker genes to viability genes controlling inbreeding depression. Theoretical and Applied Genetics **88**: 925–932 (1994a). — FU, Y. B. and RITLAND, K.: Marker-based inferences about fecundity genes contributing to inbreeding depression in *Mimulus guttatus*. Genome **37**: 1005–1010 (1994b). — HEDRICK, P. W. and MUONA, O.: Linkage of viability genes to marker loci in selfing organisms. Heredity **64**: 67–72 (1990). — HILL, A. P.: Quantitative linkage: a statistical procedure for its detection and estimation. Annals of Human Genetics **38**: 439–449 (1975). — HUSBAND, B. C. and SCHEMSKE, D. W.: Evolution of the magnitude and timing of inbreeding depression in plants. Evolution **50**: 54–70 (1996). — JAYAKAR, S. D.: On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. Biometrics **26**: 451–464 (1970). — KEMPTHORNE, O.: An introduction to genetic statistics. Iowa State University Press, Ames, IA, USA. 545 p. (1973). — KNOTT, S. A., ELSEN, J. M. and HALEY, C. S.: Methods for multiple marker mapping of quantitative trait loci in half-sib populations. Theoretical and Applied Genetics **93**: 71–80 (1996). — KUANG, H., RICHARDSON, T. E., CARSON, S. D. and BONGARTEN, B. C.: An allele responsible for seedling death in *Pinus radiata* D. DON. Theoretical and Applied Genetics **96**: 640–644 (1998). — KUMAR, S., CARSON, S. D. and GARRICK, D. J.: Detecting linkage between a fully-informative marker locus and a trait locus in outbred populations using analysis of variance. (Submitted to Forest Genetics) (1998). — LANDER, E. S. and BOTSTEIN, D.: Mapping MENDELian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185–199 (1989). — LINDGREN, D.: Use of selfed material in forest tree improvement. Department of Forest Genetics, Royal College of Forestry, Research Notes No. 15. 76 p. Stockholm, Sweden (1975). — LUO, Z. W.: 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 (1993). — MUONA, O., YAZDANI, R. and RUDIN, D.: Genetic changes between life stages in *Pinus sylvestris:* Allozyme variation in seeds and planted seedlings. Silvae Genetica **36**: 39–42 (1987). — MURANTY, H.: Power of tests for quantitative trait loci detection using full-sib families in different schemes. Heredity **76**: 156–165 (1996). — PLESSAS, M. E. and STRAUSS, S. H.: Allozyme differentiation among populations, stands and cohorts in Monterey pine. Canadian Journal of Forest Research **16**: 1155–1164 (1986). — PLOMION, C., DUREL, C.-E. and O'MALLEY, D. M.: Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. Theoretical and Applied Genetics **93**: 849–858 (1996). — SCHAFER-PREGL, R., SALAMINI, S. and GEBHARDT, C.: 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 (1996). — SKRØPPA, T.: Diallel crosses in *Picea abies*. II. Performance and inbreeding depression of selfed families. Forest Genetics **3**: 69–79 (1996). — SOLLER, M. and GENIZI, A.: 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 (1978). — WELLER, J. I., KASHI, Y. and SOLLER, M.: 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 (1990). — WILCOX, M. D.: Inbreeding depression and genetic variances estimated from self- and cross-pollinated families of *Pinus radiata*. Silvae Genetica **32**: 89–96 (1983). — WILLIAMS, C. G. and NEALE, D. B.: Conifer wood quality and marker-assisted selection: A case study. Canadian Journal of Forest Research **22**: 1009–1017 (1992). — WILLIAMS, C. G. and SAVOLAINEN, O.: Inbreeding depression in conifers: Implications for breeding. Forest Science **42**: 102–117 (1996). — YADZANI, R., MUONA, O., RUDIN, D. and SZMIDT, A. E.: Genetic structure of a *Pinus sylvestris* L. seed tree stand and naturally regenerated under-story. Forest Science **31**: 430–436 (1985).