Theor Appl Genet (2004) 108:292–298
DOI 10.1007/s00122-003-1352-7

ORIGINAL PAPER

**S. Kumar · C. Echt · P. L. Wilcox · T. E. Richardson**

# Testing for linkage disequilibrium in the New Zealand radiata pine breeding population

Received: 13 September 2002 / Accepted: 29 April 2003 / Published online: 30 July 2003


**Abstract** Linkage analysis is commonly used to find marker-trait associations within the full-sib families of forest tree and other species. Study of marker-trait associations at the population level is termed linkage-disequilibrium (LD) mapping. A female-tester design comprising 200 full-sib families generated by crossing 40 pollen parents with five female parents was used to assess the relationship between the marker-allele frequency classes obtained from parental genotypes at SSR marker loci and the full-sib family performance (average predicted breeding value of two parents) in radiata pine (*Pinus radiata* D. Don). For alleles (at a marker locus) that showed significant association, the copy number of that allele in the parents was significantly correlated, either positively or negatively, with the full-sib family performance for various economic traits. Regression of parental breeding value on its genotype at marker loci revealed that most of the markers that showed significant association with full-sib family performance were not significantly associated with the parental breeding values. This suggests that over-representation of the female parents in our sample of 200 full-sib families could have biased the process of detecting marker-trait associations. The evidence for the existence of marker-trait LD in the population studied is rather weak and would require further testing. The exact test for genotypic disequilibrium between pairs of linked or unlinked marker loci revealed non-significant LD. Observed genotypic frequencies at several marker loci were significantly different from the expected Hardy-Weinberg equilibrium. The possibilities of utilising marker-trait associations for early selection, among-family selection and selecting parents for the next generation of breeding are also discussed.

Communicated by D.B. Neale

S. Kumar (✉) · C. Echt · P. L. Wilcox · T. E. Richardson
N Z Forest Research Institute, Private bag 3020, Rotorua, New Zealand,
e-mail: Satish.Kumar@forestresearch.co.nz
Tel.: +64-7343-5899
Fax: +64-7343-5330

**Keywords** Linkage disequilibrium · SSR marker · Breeding value · Marker-assisted selection · *Pinus radiata*

## Introduction

The availability of genetic markers has facilitated experimental studies to detect quantitative trait loci (QTLs) in many species. Loci affecting quantitative traits can be detected via their linkage with these markers, only if there is linkage disequilibrium (LD) between the marker and the actual QTL. However, in most outcrossed forest-tree populations the association between alleles at a marker locus and the true QTL is likely to be in linkage equilibrium (Strauss et al. 1992). To overcome this limitation, most QTL detection experiments in the forest trees use family structures (e.g. full-sibling detection populations) that generate LD. Using this approach, a number of studies have found putative QTLs influencing various economic traits in conifers (e.g. Knott et al. 1997; Kumar et al. 2000; Sewell et al. 2001).

Simulation studies (e.g. Johnson et al. 2000; Kumar and Garrick 2001) and case studies (e.g. Williams and Neale 1992) have shown that substantial additional genetic gain can be obtained using within-family marker-assisted selection (MAS) in conifers. There are two major areas where MAS could contribute to such gains, namely 'backward' selection (re-selection of parents) and 'forward' selection (selection of individuals within families), in tree-breeding programs. For successful integration of MAS for 'backward' selection and/or 'forward' selection, it is important that marker-QTL associations are stable across a variety of genetic backgrounds. However, high degrees of linkage equilibrium (which means that linkage-phase relationships differ among individuals) makes broader application of marker-assisted selection (MAS) in forest trees very complex. This characteristic generally precludes the use of MAS for across-family selection and requires that MAS for within-family selection be customised for each family.

Using a factorial mating design, Arcade et al. (1996) provided experimental evidence for the existence of marker-trait associations in a broader population of larch. Verhaegen et al. (1998) using a factorial design in *Eucalyptus*, found a significant relationship between the cumulative number of marker alleles in the parents with the full-sib family performance for various traits. Sewell et al. (2001) showed that marker-trait associations were stable across two related pedigrees of loblolly pine. Finding molecular markers such as these will enhance the rate of genetic gain by using such markers in the between-family and within-family selections (e.g. Neale et al. 2002).

Natural and artificial selection, genetic sampling effects due to small effective population sizes, hybridisation and restricted migration of gametes among subdivisions, can produce LD among loci. In addition, LD might arise from selection on epistatic interaction (Lewontin 1974) or genetic drift (Hill and Robertson 1968) caused by the geographic origin of the populations involved in the breeding program. Verhaegen et al. (1998), in a LD study of *Eucalyptus* hybrids, postulated the different natural origins of the populations involved as the cause of the LD in their study. The current breeding population of New Zealand radiata pine has originated from local 'land-race' stocks which in turn have resulted essentially from hybridisation between two distinct natural populations (Monterey and Año Nuevo).

Non-random mating and artificial selection over 3–4 generations have led to the current radiata-pine breeding population. It is expected that provenance-hybridisation, non-random mating and artificial selection could have caused LD in the current New Zealand radiata-pine breeding population. Availability of linkage maps (e.g. Devey et al. 1999; Sewell et al. 1999) constructed using highly polymorphic markers provides an opportunity to test if LD exists between alleles at linked loci in breeding populations that have gone through generations of artificial selection and breeding.

Testing the LD hypothesis by detecting marker-QTL associations in each of the families in a large breeding program, such as in New Zealand, can be prohibitively costly. In this study, we extend the approach, used by Arcade et al. (1996) and Verhaegen et al. (1998) for dominant markers (RAPD) to codominant markers, to test the existence of LD at the population level between a series of marker loci and trait loci for various growth and form traits. The other two objectives of this study were to test for genotypic disequilibrium between pairs of various linked and unlinked marker loci; and to test the existence of disequilibrium at a given marker locus [this is similar to testing for departure from Hardy-Weinberg (H-W) equilibrium].

## Materials and methods

### Genetic material and female-tester mating design

The concept of "female-testing" assumes that the tested progeny of crosses of candidate pollen parents with four or five females give an accurate estimate of the breeding value or the general combining ability (GCA) of the pollen parents. A field trial consisting of 750 pair-crosses (150 pollen parents mated factorially with each of five female parents) was established in 1994 with six replicates of each pair cross. Each tree was assessed at the age of 6-years for diameter at breast height (mm), stem straightness (1–9 scale; 1 = most crooked, 9 = very straight), branching cluster frequency (1–9 scale; 1 = fewest branch clusters, 9 = most clusters) and wood density (kg/$m^3$).

To calculate variance components and to predict the parental breeding values (which will be used to study marker-trait association), the following linear model was used:

$$Y_{ijkl} = \mu + R_i + P_j + T_k + PT_{jk} + e_{ijkl}, \quad (1)$$

where $Y_{ijkl}$ is the observation for the $l^{\text{th}}$ progeny of the cross between $k^{\text{th}}$ tester and $j^{\text{th}}$ pollen parent in the $i^{\text{th}}$ replicate; $\mu$ is an overall mean; $R_i$ is the fixed effect associated with the $i^{\text{th}}$ replicate; $P_j$ is the random effect associated with the $j^{\text{th}}$ pollen parent; $T_k$ is the random effect associated with the $k^{\text{th}}$ female tester; $PT_{jk}$ is the random interaction associated with the $j^{\text{th}}$ pollen parent and $k^{\text{th}}$ female tester; and $e_{ijkl}$ is the random error.

The random effects, $P_j$, $T_k$, $PT_{jk}$ and $e_{ijk}$, are assumed to have expected values equal to zero and variances $\sigma^2_c$, $\sigma^2_t$, $\sigma^2_{ct}$ and $\sigma^2_e$, respectively. ASREML software (Gilmour et al. 1997) was used to estimate genetic parameters and to obtain the best linear unbiased prediction (BLUP) of parental breeding values.

### Molecular-marker assays

For the purpose of this linkage disequilibrium study, genotyping of the 45 parents (40 pollen parents and 5 female parents) using 34 polymorphic SSR markers was out-sourced to ***SignaGen***—Molecular Breeding Solutions (http://www.SignaGen.com). All individuals produced reliable 2-allele genotypes at each locus. More detail of the methodology is available on request from ***SignaGen***. Biallelic genotypes at each of the 34 loci were obtained for each of the 45 parents and the allele frequencies were obtained.

As shown in Table 1, frequencies of observed alleles varied from 0.01 to 0.82. Most of the alleles were present in very low frequency. The three highest-frequency alleles were first identified and the remaining low-frequency alleles were lumped to form a "synthetic" 4th allele. This resulted in four alleles (three distinct alleles and one synthetic allele) at each locus. If there were originally only four observed alleles at a locus, and one of the alleles had a frequency less than 5%, then it was lumped with the 3rd most-frequent allele and thus resulted in three alleles (two distinct alleles and one synthetic allele) at all such loci. Synthetic alleles were used because if the true rare allele was present only in female parents, it may bias the process of detecting marker-trait association in the breeding population using the female-tester design (Ewens and Spielman 1995; Pritchard and Rosenberg 1999). Biallelic genotypes of all 45 parents were obtained after forming the "synthetic" alleles at each marker locus and were used for all further calculations.

### LD between a marker and a trait locus

Analysis of variance (ANOVA) is commonly used to study the association between marker genotypes and trait loci in segregating populations. Arcade et al. (1996) and Verhaegen et al. (1998) used one-way ANOVA to investigate the statistical relationship between the expected RAPD marker-"band present" allele-frequency in the offspring, and family performance. The expected frequency of the

**Table 1** Number of alleles and their frequencies (calculated from a sample size of 45 individuals) at each marker loci. The linkage groups to which a marker was assigned to are based on a QTL mapping pedigree

| Locus abbreviation | No. of alleles | Range of allele frequencies | No. of alleles after lumping | Linkage group |
|---|---|---|---|---|
| NZPR0119 | 9 | 0.01–0.34 | 4 | 15 |
| NZPR0129 | 21 | 0.01–0.24 | 4 | 6 |
| NZPR0138b | 14 | 0.01–0.23 | 4 | 3 |
| NZPR0143 | 9 | 0.02–0.25 | 4 | 3 |
| NZPR0157 | 12 | 0.02–0.23 | 4 | 1 |
| NZPR0248b | 10 | 0.01–0.35 | 4 | 15 |
| NZPR0290 | 23 | 0.01–0.15 | 4 | 1 |
| NZPR0407 | 15 | 0.01–0.27 | 4 | 11 |
| NZPR0413 | 14 | 0.01–0.25 | 4 | 4 |
| NZPR0440 | 4 | 0.03–0.73 | 3 | 11 |
| NZPR0443 | 8 | 0.01–0.53 | 4 | 7 |
| NZPR0476 | 7 | 0.01–0.82 | 3 | 16 |
| NZPR0495 | 9 | 0.01–0.29 | 4 | 11 |
| NZPR0581 | 7 | 0.01–0.35 | 4 | 10 |
| NZPR0599 | 10 | 0.01–0.19 | 4 | 16 |
| NZPR0706 | 7 | 0.01–0.37 | 4 | 1 |
| NZPR0826 | 6 | 0.01–0.56 | 4 | 6 |
| NZPR0947 | 3 | 0.08–0.70 | 3 | 15 |
| NZPR1078 | 4 | 0.01–0.76 | 4 | 3 |
| PR4.6 | 14 | 0.01–0.32 | 4 | 1 |
| RIPPT0079a | 6 | 0.01–0.35 | 4 | 7 |
| RIPPT0699 | 18 | 0.01–0.40 | 4 | 1 |
| RIPPT0914 | 7 | 0.02–0.61 | 4 | 2 |
| RIPPT0984 | 22 | 0.01–0.20 | 4 | 10 |
| UCC0705 | 4 | 0.03–0.48 | 3 | 7 |
| UCC0719 | 5 | 0.01–0.69 | 3 | 16 |
| UCC0735 | 8 | 0.01–0.60 | 4 | 2 |
| UCC0759 | 16 | 0.01–0.24 | 4 | 6 |
| UCC0784 | 10 | 0.01–0.45 | 4 | 8 |
| UCC0802 | 14 | 0.01–0.34 | 4 | 3 |
| UCC0844 | 11 | 0.01–0.32 | 4 | 3 |
| UCC0922 | 14 | 0.01–0.39 | 4 | 6 |
| UCC1017 | 15 | 0.01–0.21 | 4 | 4 |
| UCC1149 | 9 | 0.04–0.35 | 4 | 6 |

"band present" allele-frequency in the offspring was obtained from the parental genotypes of the family. Because RAPDs are dominant markers, they grouped full-sib families into either two or three frequency classes, depending on whether the marker was polymorphic in one or both parents. Using codominant markers in the present study, it was possible to group full-sib families in up to five expected frequency classes of an allele, depending upon the marker genotypes of the parents (see Fig. 1).

Lande and Thompson (1990) suggested the use of the regression of phenotypic records of the quantitative trait on the number of favourable marker alleles at a marker locus to infer the presence of linkage disequilibrium. Luo (1998), using simulation studies, compared the power of ANOVA and regression approaches, and reported that regression analysis is more powerful than the ANOVA, particularly when the QTL has low heritability. In this study, the concept of Arcade et al. (1996) and Verhaegen et al. (1998) was fitted using a linear regression approach. For each of the alleles at 34 marker loci, the linear model was written as:

$$F_i = \mu + \gamma_{jk} + e_{ijk}, \quad (2)$$

where $F_i$ is the expected performance (defined as the average breeding value of the two parents; breeding value estimates were obtained from the female-tester design as explained earlier) of $i^{th}$ full-sib family; $\mu$ is the general mean; $\gamma_{jk}$ is the expected frequency (calculated from the parental marker genotypes) of allele $j$ at the $k^{th}$ locus in the full-sib progeny; $e_{ijk}$ is the random error. A total of 524 regression analyses were conducted for various combinations of alleles at different marker loci and trait phenotypes. The significance of the regression coefficient was used to infer LD at the population level. As the mating design structure involved only five female parents, some spurious associations between marker and the pair-cross effects are expected. Thus, the analysis of regression of parental breeding values or general combining ability (GCA) on marker allele-frequency was also undertaken to infer LD in the breeding population. The sample sizes within each of the allele-frequency class could vary and thus to account for this unbalanced structure, the threshold levels for testing the significance of the regression coefficient was obtained using the permutation method (Churchill and Doerge 1994).

## LD between pairs of marker loci

Genotypic data on pairs of loci was used to calculate the LD between pairs of marker loci. When working with genotypic data from natural populations, it is not reasonable to assume that random-mating occurs, and gamete frequencies are seldom known. Hill (1974) proposed a method to calculate LD coefficients when the population undergoes the random union of gametes. As some of the single-locus results for our population indicate departures from H-W frequencies (see below), an assumption that the random union of gametes can not be assumed. Weir (1979) proposed a composite measure of LD, which is appropriate under conditions of the non-random union of gametes and incomplete identification of genotypes. To test the genotypic LD between a pair of loci, the null hypothesis that the genotypes at one locus are independent from genotypes at the other locus was assumed.

The analysis was implemented using GENEPOP software (Raymond and Rousset 1995). Using this method, a contingency table based on the observed genotypic data, is prepared for each pair of loci and an unbiased estimate of the *P*-value of the probability test, or the exact test is performed using a Markov Chain method. The *P*-value of the test is calculated as the sum of the probabilities of all contingency tables (with the same marginal

values as the observed one) with a lower or equal probability than the observed table. Acceptance or rejection of the null hypothesis is based on a comparison of the *P*-value with a preset $\alpha$ level (Louis and Dempster 1987).

### Test for H-W equilibrium at each marker locus

Genotypic data from 45 individuals was used to test the null hypothesis that the gametes are united at random. This analysis was implemented using GENEPOP software (Raymond and Rousset 1995). Unbiased estimates of the *P*-value along with their standard-error estimate of the exact test (Haldane 1954) was obtained for each of the 34 marker loci. The global test across loci was implemented using the multi-sample score test (Rousset and Raymond 1995) in GENEPOP. An estimate of the correlation of genes within individuals within a population ($F_{IS}$, Weir and Cockerham 1984), at each marker locus, was also obtained.

## Results and discussion

### H-W equilibrium at each marker locus

The unbiased estimate of the *P*-value of the exact test along with the standard error at each marker locus is given in Table 2. At seven marker loci, the observed genotypic frequencies were significantly (at the 5% comparison-wise threshold level) different from the expected H-W frequencies (Table 2). These results support the notion of the non-random union of gametes at some loci. However, after applying Bonferroni correction to account for multiple testing across 34 loci, there was non-significant departure from H-W frequencies at all 34 loci. When a more-specific alternative hypothesis of heterozygote excess was tested, the null hypothesis of random union of gametes was accepted at all 34 marker loci (data not shown). Global deviation from H-W equilibrium was also found to be insignificant.

The null hypothesis was rejected (at the 5% comparison-wise threshold level) in favour of the alternative hypothesis of heterozygote deficit (or high $F_{IS}$) at seven (NZPR0138b, NZPR0440, NZPR0476, NZPR0495, NZPR0581, UCC0784 and UCC1149) marker loci (data not shown). However, after applying the Bonferroni correction, only two loci (NZPR0138b and NZPR0581) remained significant at the 5% experiment-wise threshold. One possible reason for apparent heterozygote deficiency could be the high inbreeding coefficient. Because the parental individuals considered in this study represents plus-trees selected from stands planted in the 1920s and 1930s, it is highly unlikely that there are inbred individuals involved. A frequent cause of apparent heterozygote deficiency with microsatellites is the failure of amplification of some alleles in the presence of some other alleles. Both of these marker loci (NZPR0138b and NZPR0581) were tested on two unrelated full-sib pedigrees and we found no indication of null alleles. Assortative mating and artificial selection over several generations might have caused departure of the genotypic frequencies from that expected under H-W equilibrium. It is also likely to detect two (=0.05 × 34) significant values due to chance alone at a probability level of 0.05.

**Table 2** Unbiased estimate of *P*-value with the standard error (SE) of the exact test for H-W equilibrium at each marker locus. The estimate of correlation of genes within individuals ($F_{IS}$) is also given

| Locus | *P*-value | SE | $F_{IS}$ |
|---|---|---|---|
| NZPR0157 | 0.4539 | 0.0022 | 0.165 |
| NZPR0290 | 0.9781 | 0.0004 | –0.034 |
| NZPR0440 | 0.0671 | 0.0010 | 0.273 |
| NZPR0476 | 0.0197 | 0.0005 | 0.289 |
| NZPR0581 | 0.0265 | 0.0007 | 0.293 |
| NZPR0706 | 0.1206 | 0.0014 | –0.115 |
| NZPR0826 | 0.4308 | 0.0022 | –0.120 |
| NZPR0947 | 0.2239 | 0.0015 | 0.207 |
| NZPR1078 | 0.9024 | 0.0008 | –0.033 |
| PR4.6 | 0.1285 | 0.0015 | 0.017 |
| RIPPT0079a | 1.0000 | 0 | 0.042 |
| RIPPT0914 | 0.7352 | 0.0018 | –0.111 |
| UCC0705 | 0.4724 | 0.0017 | 0.117 |
| UCC0719 | 0.5856 | 0.0015 | 0.072 |
| UCC0735 | 0.2488 | 0.0020 | –0.205 |
| UCC0802 | 0.0313 | 0.0008 | –0.063 |
| UCC0844 | 0.8663 | 0.0011 | 0.026 |
| NZPR0119 | 0.7504 | 0.0017 | –0.090 |
| NZPR0129 | 0.0293 | 0.0008 | 0.094 |
| NZPR0138b | 0.0057 | 0.0003 | 0.400 |
| NZPR0143 | 0.6257 | 0.0019 | 0.135 |
| NZPR0248b | 0.9612 | 0.0005 | –0.009 |
| NZPR0407 | 0.6544 | 0.0020 | 0.0190 |
| NZPR0413 | 0.0278 | 0.0007 | –0.100 |
| NZPR0443 | 0.3560 | 0.0022 | –0.108 |
| NZPR0495 | 0.1030 | 0.0012 | 0.186 |
| NZPR0599 | 0.1875 | 0.0018 | 0.100 |
| RIPPT0699 | 0.9626 | 0.0005 | –0.005 |
| RIPPT0984 | 0.4304 | 0.0024 | 0.093 |
| UCC0759 | 0.2430 | 0.0020 | 0.060 |
| UCC0784 | 0.0282 | 0.0007 | 0.217 |
| UCC0922 | 0.9458 | 0.0006 | 0.007 |
| UCC1017 | 0.1024 | 0.0013 | –0.077 |
| UCC1149 | 0.1101 | 0.0013 | 0.229 |

### LD between pairs of marker loci

Genotypic disequilibrium between all possible pairs of 34 marker loci was obtained. The unbiased estimates of *P*-value for all pairs of loci are too numerous to report here. There were 50 significant (at the 5% comparison-wise threshold level) pairs out of 561 possible pairs. Two pairs of markers mapped on linkage group 15 (NZPR0119 and NZPR0248b) and linkage group 1 (NZPR0157 and PIPPT0699) showed significant genotypic disequilibrium (data not shown). Some pairs of marker loci that were mapped to different linkage groups (LGs) also showed significant genotypic disequilibrium between them. For example, PR4.6 and UCC0719 showed a significant LD even though these loci were mapped on different LGs. While PR4.6 was found to be associated with wood density in a QTL-mapping pedigree (Wilcox et al. 1998), the UCC0719 showed a significant association with the wood density in this study. After applying the Bonferroni

**Table 3** Summary statistics, at the individual-tree level, and estimates of narrow-sense heritability ($h^2$) of various traits: diameter at breast height (DBH), straightness (STR), branch cluster frequency (BR) and wood density (WD)

| Statistics | DBH (mm) | STR (1–9 scale) | BR (1–9 scale) | WD (kg/m$^3$) |
|---|---|---|---|---|
| Mean | 174 | 6.11 | 5.76 | 353 |
| Minimum | 63 | 1 | 1 | 271 |
| Maximum | 300 | 9 | 9 | 439 |
| $h^2$ | 0.14 | 0.17 | 0.20 | 0.68 |

**Table 4** Significant markers and the variance explained for different traits

| Trait | Marker | $R^2$ (%) for the positive allele | $R^2$ (%) for the negative allele |
|---|---|---|---|
| DBH | NZPR0440 | 30.0 | 47.8 |
| | UCC0784 | 25.4 | 23.6 |
| BR | NZPR0143 | 41.8 | 46.3 |
| | NZPR0138b | 42.8 | 33.6 |
| STR | NZPR0290 | 21.5 | 24.8 |
| | NZPR0706 | 19.7 | 14.6 |
| WD | UCC0719 | 9.7 | 14.20 |

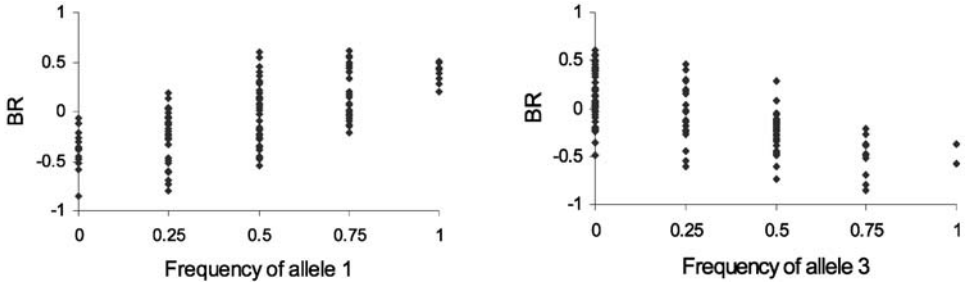


**Fig. 1** Association of the alleles 1 and 3 with the family performance for branching (*BR*) at the marker locus NZPR0143. The other alleles at this locus did not show any association. Expected frequency of the allele in the full-sib families is shown on the X-axis and the expected family performance is shown on the Y-axis

correction, none of the LD tests between marker loci remained significant. As the majority of the marker loci were mapped with a recombination fraction of more than 0.05, one would not expect to detect strong genotypic disequilibrium among various pairs of loci.

## LD between a marker and a trait locus

Marker-trait associations were evaluated for various traits measured at 6-years of age: diameter at breast height (DBH), stem straightness (STR), branching cluster frequency (BR) and wood density (WD). Summary statistics along with estimates of narrow-sense heritability of various traits, obtained from complete trial data, are shown in Table 3. At the individual-tree level, the average DBH, STR, BR and WD were 174, 6.11, 5.76 and 353 respectively. Estimates of narrow-sense heritabilities were 0.14, 0.17, 0.20 and 0.68 for DBH, STR, BR and WD respectively.

Molecular markers that showed significant (at the 1% experiment-wise threshold level) associations with the full-sib family performance for a number of traits, are shown in Table 4. In such cases, at each marker locus, the alleles showing positive, negative or no effect on the family performance were clearly evident. For example, Fig. 1 shows that at locus NZPR0143 the average family performance for BR increases as the expected frequency of allele 1 increases in the progeny. The average family performance for the BR score was found to be –0.38, –0.25, 0.003, 0.019 and 0.38 when the expected frequency of allele 1 in the families was 0, 0.25, 0.50, 0.75 and 1.0, respectively. At locus NZPR0143 allele 3 shows an association with a negative QTL allele for branching. For allele frequency of 0, 0.25, 0.50, 0.75 and 1.0, the expected family performance was 0.18, –0.05, –0.25, –0.51 and –0.48 (Fig. 1). The molecular marker (NZPR0143) showing association with the BR was mapped in the same region that showed significant association with BR in a QTL-mapping pedigree (Wilcox et al. 1998). For WD, the average expected family performance increased from –5.24 to 2.91 kg/m$^3$ when the frequency of allele 3 increased from 0 to 1. Similarly, the average WD of families decreased from 2.11 to –6.17 kg/m$^3$ with a change of allele-2 frequency from 0 to 1 (Fig. not shown).

As the design used in this study involved only five female parents and 40 male parents, there is a possibility that the significant associations found at different marker loci might be influenced by the fact that these alleles are

**Table 5** Estimated frequency of the favourable and unfavourable alleles in male and female populations, at the selected marker loci

| Locus | Frequency of the positive-effect allele in | | Frequency of the negative-effect allele in | |
|---|---|---|---|---|
| | Males | Females | Males | Females |
| UCC0784 | 0.22 | 0.30 | 0.18 | 0.40 |
| NZPR0440 | 0.14 | 0.40 | 0.76 | 0.50 |
| NZPR0143 | 0.41 | 0.50 | 0.20 | 0.33 |
| NZPR0138b | 0.44 | 0.36 | 0.24 | 0.17 |
| NZPR0290 | 0.14 | 0.25 | 0.64 | 0.75 |
| NZPR0706 | 0.35 | 0.50 | 0.21 | 0.10 |
| UCC0719 | 0.16 | 0.50 | 0.71 | 0.50 |

segregating only in the female population. Any marker allele that is in high frequency in the over-represented sub-population (i.e. five female testers in our case) would show association with the phenotype (Ewens and Spielman 1995; Pritchard and Rosenberg 1999). To study this aspect, the frequency of the alleles showing significant association with full-sib family performance, was calculated for male and female populations separately. Results shown in Table 5 indicated that the frequency of favourable alleles was generally higher in the female population, but no such trend was observed for unfavourable alleles. These results reflected that the evidence of LD found using full-sib family performance could have been influenced by the mating-design structure.

To resolve this issue further, we investigated the relationship between the predicted breeding value (or GCA) of a parent to its genotype at the marker loci that showed a significant LD with the full-sib family performance. Most of the markers showing significant association with the full-sib family performance in the initial analysis (see Table 4) were found to show non-significant associations at the parental GCA level. However, allele 3 and allele 2 at the marker loci NZPR0143 and UCC0719, respectively, were significantly ($P$ = 0.0004 and 0.0541, respectively) associated with the predicted breeding values for BR and WD, respectively. After applying Bonferroni correction, allele 3 at the marker loci NZPR0143 remained significant at 20% experiment-wise threshold.

It is interesting to note that all marker-trait associations found using full-sib family performance could not be repeated when the parental GCA was regressed on the marker-allele frequencies. There could be various explanations for these results. The number of significant tests found (7 out of 524) were not much different from that would be expected (5 out of 524) due to chance alone at a probability level of 0.01. Thus, most of the significant marker-trait associations could be due to the genetic-sampling effects. Mating design structure could also have contributed to these results. Over-representation of five female parents in our sample of 200 full-sib families might have contributed to the observed significant marker-trait associations (e.g. Ewens and Spielman 1995; Pritchard and Rosenberg 1999). These results clearly indicate that the existence of LD, found using full-sib performance (as was the case in Arcade et al. 1996 and Verhaegen et al. 1998), could be verified using GCA effects of parents involved in those full-sib crosses. We suggest this should be a statistical validation of the association found at the full-sib family level. Our results suggest that the evidence of marker-trait LD, in the population of 45 parents that we studied, is rather weak. Various candidate genes influencing wood property traits, along with the markers showing putative associations with trait loci in this pilot study, are now being evaluated using a larger set of the New Zealand radiata pine-breeding population parents.

## Marker-assisted selection for breeding

Various authors have advocated the use of molecular markers in advanced generation tree-breeding programmes (e.g. O'Malley and McKeand 1994; Johnson et al. 2000; Kumar and Garrick 2001). A prominent feature of advanced generation programmes is the between-and-within-family selection, where full-sib families are culled based on the mid-parent GCA values and the best offspring are chosen for breeding from the remaining families. In the New Zealand radiata pine-breeding programme, various new elite (or breed) populations are being formed for various economic traits (e.g. structural and appearance timber grades) by selecting 24 parents from the main breeding population (Jayawickrama and Carson 2000). These parents are crossed factorially to create full-sib families where advanced generation selections are made for breeding and deployment purposes. Marker-trait associations found using full-sib family performance and parental GCA values, as described in this study, would indicate the QTLs of 'general value' that can be used for between-and-within-family selection in elite populations. Detecting LD in small elite populations, however, might not be generalised for the wider radiata pine populations. The method described here would assist finding QTLs that are specific to the best existing parental combinations. The most-heralded benefit of MAS in forestry is the use of marker-trait associations for early selection. The marker-trait LD found using the approach presented in this study could be used for early selection of individuals for mass propagation or for clonal testing.

**Acknowledgements** We thank the New Zealand Radiata Pine Breeding Company for giving access to phenotypic data. We thank Dr. Christophe Plomion and Dr. Rowland Burdon for helpful

discussion during the completion of this project. We also thank anonymous referees for very helpful and constructive comments. The experiments comply with the current laws of New Zealand where the experiments were performed. SSR markers with prefixes RIPPT and UCC are exclusive property of International Paper Company. We thank them for the use of these markers for this study.

## References

Arcade A, Faivre-Rampant P, Le Guerroue B, Paques LE, Prat D (1996) Quantitative traits and genetic markers: analysis of a factorial mating design in larch. In: Ahuja MR, Boerjan W, Neale D (eds) Somatic cell genetics and molecular genetics of trees. Kluwer Academic Publishers, Doedrecht, The Netherlands, pp 211–216

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Devey ME, Sewell MM, Uren TL, Neale DB (1999) Comparative mapping in loblolly pine and radiata pine using RFLP and microsatellite markers. Theor Appl Genet 99:656–652

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet 57:455–464

Gilmour AR, Thompson R, Cullis BR, Welham SJ (1997) ASREML. NSW Agriculture, Orange, NSW, Australia

Haldane JBS (1954) An exact test for randomness of mating. J Genet 52:631–635

Hill WG (1974) Estimation of linkage disequilibrium in random mating populations. Heredity 33:229–239

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Jayawickrama, KJS, Carson MJ (2000) A breeding strategy for the New Zealand Radiata Pine Breeding Cooperative. Silvae Genet 49:82–90

Johnson GR, Wheeler NC, Strauss SH (2000) Financial feasibility of marker-aided selection in Douglas-fir. Can J For Res 30:1942–1952

Knott SA, Neale DB, Sewell MM, Haley CS (1997) Multiple-marker mapping of quantitative trait loci in an outbred pedigree of loblolly pine. Theor Appl Genet 94:810–820

Kumar S, Garrick DJ (2001) Genetic response to within-family selection using molecular markers in some radiata pine breeding schemes. Can J For Res 31:779–785

Kumar S, Spelman RJ, Garrick DJ, Richardson TE, Lausberg M, Wilcox PL (2000) Multiple-marker mapping of wood density loci in an outbred pedigree of radiata pine. Theor Appl Genet 100:926–933

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lewontin RC (1974) The genetic basis of evolutionary change. Columbia Univ. Press, NewYork

Louis EJ, Dempster ER (1987) An exact test for Hardy-Weinberg and multiple alleles. Biometrics 43:805–811

Luo ZW (1998) Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. Heredity 80:198–208

Neale DB, Sewell MM, Brown GR (2002) Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. Ann For Sci 59:595–605

O'Malley DM, McKeand SE (1994) Marker-assisted selection for breeding value in forest trees. For Genet 1:207–218

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228

Raymond M, Rousset F (1995) GENEPOP: population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. Genetics 140:1413–1419

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

Sewell MM, Bassoni DL, Megraw RA, Davis, MF, Wheeler, NC, Tuskan, GA, Neale DB (2001) Mapping QTLs for wood property traits in loblolly pine (*Pinus taeda* L.) with application towards marker-assisted breeding (Abstract). Wood, Breeding and Biotechnology Workshop, Bordeaux, 11th–14th, June 2001

Strauss SH, Lande R, Namkoong G (1992) Limitations of molecular marker-aided selection in forest-tree breeding. Can J For Res 22:1050–1061

Verhaegen D, Plomion C, Poitel M, Costa P, Kremer A (1998) Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers 2. Linkage disequilibrium in a factorial design between *E. urophylla* and *E. grandis*. For Genet 5:61–69

Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35:235–254

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Wilcox PL, Richardson TE, Corbett GC, Lee JR, Lausberg M, Carson SD (1998) Results from Phase-II QTL detection experiments. New Zealand Forest Research Institute, Report No. 61

Williams CG, Neale DB (1992) Conifer wood quality and marker-aided selection: a tool for the improvement of forest-tree species. Can J For Res 22:1009–1017