Theor Appl Genet (1996) 93: 1083–1089

C. Plomion B.-H. Liu D. M. O'Malley

Genetic analysis using trans-dominant linked markers in an F_2 family

Received: 7 March 1996 / Accepted: 2 August 1996

Abstract Trans-dominant linked markers pairs (trans referring to the repulsion linkage phase) provide a model for inferring the F_2 progeny genotype based upon both the conditional probabilities of F_2 genotypes, given the F_2 phenotype, and prior information on marker arrangement. Prior information of marker arrangement can be readily obtained from a linkage analysis performed on marker segregation data in a family resulting by crossing the F_1 individual to a "tester" parent or else can be obtained directly from the gametes of the F_1, or from recombinant inbred lines. We showed that a trans-dominant linked marker (TDLM) pair can be recoded as a "co-dominant megalocus" when the recombination fraction, r_1, for a pair of TDLMs is less than 0.05. We obtained a maximum-likelihood estimator (MLE) of the recombination frequency, r_2, between a TDLM pair and a co-dominant marker in an F_2 family using the EM algorithm. The MLE was biased. Mean bias increased as r_1 and r_2 increased, and decreased as sample size increased. The information content for r_2 was compared to the information content of dominant and co-dominant markers segregating in an F_2 family. It was almost identical with two co-dominant markers when $r_1 \leq 0.01$ and $r_2 \geq 0.05$. For larger values of r_1, $(0.05 \leq r_1 \leq 0.15)$ a TDLM pair provided 75%–66% of the information content of two co-dominant markers. Although dominant markers can be converted to co-dominant markers by a laborious process of cloning, sequencing, and PCR, TDLM pairs could easily substitute for co-dominant markers in order to detect quantitative trait loci (QTLs) and estimate gene action in an F_2 family.

Key words Dominant marker · F_2 family · Linkage · QTL · Information function

Communicated by G. E. Hart

C. Plomion (✉)
INRA, Laboratoire de Génétique et Amélioration des Arbres Forestiers, BP45, F-33610 Cestas, France

B.-H. Liu D. M. O'Malley
Forest Biotechnology Group, Department of Forestry, North Carolina State University, Box s8008 Raleigh, NC 27695, USA

Introduction

Co-dominant markers, such as restriction fragment length polymorphisms (RFLPs), have often been used for quantitative trait dissection analysis in F_2 families (e.g. Edwards et al. 1987; Paterson et al. 1991a) because all of the possible gene dosages are represented and gene action can be defined. Arbitrarily primed PCR-based molecular marker systems (Welsh and McClelland 1990; Williams et al. 1990; Caetano-Anolles et al. 1991; Vos et al. 1995) have demonstrated considerable advantages over hybridization-based methods in terms of their efficiency to acquire data, little need for prior genetic information, the small amounts of DNA needed, and the shorter time required for implementation, but these markers often have a dominant mode of inheritance. Knapp et al. (1995) proposed that dominant marker-segregation data in an F_2 family could be analyzed to provide two maps for each homolog based on markers in coupling. However, the locus order for dominant markers in repulsion (markers on alternative homologs) is not readily determined, and the estimates of recombination fraction are biased (Knapp et al. 1995) and have excessively large sampling errors (Allard 1956). A disadvantage of using dominant markers for quantitative trait dissection analysis in an F_2 progeny, is that additive effects and dominance deviations for individual quantitative trait loci (QTLs) cannot be estimated separately.

Considerable efforts have been made to convert dominant markers to co-dominant markers (e.g., SCARs, sequenced characterized amplified regions, Paran and Michelmore 1993) because co-dominant markers have advantages for genetic analysis in F_2 populations. Nevertheless, the process required to make dominant markers co-dominant is laborious and lessens the advantages of random amplified polymorphic DNA (RAPD) markers compared with other systems. Alternatively, Williams et al. (1993) proposed that a pair of dominant markers tightly linked in repulsion, provided almost the same information as a single co-dominant marker, but a formal analysis of this "megalocus" approach has not yet been reported. In the present

paper, we describe the theoretical basis for using trans-dominant linked marker (TDLM) pairs for further QTL analysis in an F_2 family obtained by selfing an F_1 individual. We show that TDLM pairs provide a model for inferring the F_2 progeny genotype based upon (1) the conditional probabilities of F_2 genotypes, given the F_2 phenotype, and (2) prior information on marker arrangement. Prior information of dominant marker arrangement (in either coupling or repulsion, as well as map distance) for complex trait dissection in an F_2 family can be readily obtained by crossing the F_1 individual to a "tester" parent (e.g. a distantly related individual), from doubled haploids or conifer megagametophytes, or from recombinant inbred lines (Chaparro et al. 1994).

Genetic model

Consider two linked genetic markers, A and B, that are dominant and in the trans configuration (i.e. repulsion phase) in an F_1 individual. The recombination fraction for the interval A-B is r_1. In an F_2 family, this genetic model is analogous to a single co-dominant marker [megalocus (AB)] when $r_1=0$, because, by definition, A and B map to the same locus. We called this model TDLM for trans-dominant linked markers. For marker loci A and B, several different genotypes can have the same F_2 phenotype. For an F_2 progeny phenotype, k, the conditional probability, $P_i|P_k$, of each genotype, i, can be specified given the recombination fraction, r_1 (see Table 1). The expected frequency of each genotype, i, was obtained from the expected gamete frequencies [i.e. the frequency of parental gametes Ab and aB=$(1-r_1)/2$ and the frequency of recombinant gametes AB and ab=$r_1/2$].

Consider the case where the TDLM pair, (AB), is linked to a co-dominant marker, C. Because r_1 is negligibly small, we assume that C is located outside of (AB), and that no double crossovers occurred in the region A-B-C. The recombination fraction for the interval (AB)-C is r_2. We assume that recombination is homogeneous for both pollen and seed parents. The information function, $I(r_2|r_1)$, for the calculation of the standard error of the maximum-likelihood estimator of recombination frequency r_2 between (AB) and C was obtained from the expected TDLM-marker genotypes (see Table 3). The conditional probability of the occurrence of a recombination for a given genotype was calculated by treating the TDLM as a "co-dominant megalocus" (see Table 3).

Results

Expected accuracy of treating a TDLM as a "co-dominant megalocus"

Several different F_2 genotypes can have the same phenotype for a TDLM pair in an F_2 family (Table 1), but one genotype has a much higher probability than all others for every F_2 phenotypic class. Thus, if we surmised that the most probable genotype (the parental one) accounted for each phenotype, we would be correct in almost every case when the true value of r_1 is small. We plotted the conditional probabilities of the F_2 parental TDLM genotype to show the expected proportion of correctly assigned TDLM genotypes (Fig. 1). The conditional probabilities showed that for a small r_1, most (AB) phenotypes have a heterozygous genotype, AabB with an approximate frequency of 0.5, while most (Ab) and (aB) phenotypes have a homozygous genotype, Aabb and aaBB, respectively, with approximate frequencies of 0.25. The probability of F_2 parental genotype occurrence was slightly greater for homozygous than for heterozygous genotypes. The (ab) phenotype had one underlying genotype, aabb, and was the only unambiguous class; but the probability of observing this genotype is very small. Thus, a TDLM pair could be considered as a "co-dominant megalocus" when r_1 was small

Table 1 Conditional probabilities of F_2 genotypes, i, given the F_2 phenotype, k, for a trans-dominant linked marker (TDLM) pair. A and B are dominant alleles, a and b are recessive alleles, and r_1 is the recombination frequency between A and B. R and P are recombinant and parental genotypes, respectively

F_2 progeny phenotype (k)	F_2 genotype (i)	Type	Expected genotype frequency P_i	Conditional probability $P_i\|P_k$
(AB)	AABB	R	$0.25(r_1)^2$	$(r_1)^2/(2+r_1^2)$
	AABb	R	$0.5r_1(1-r_1)$	$2r_1(1-r_1^2)/(2+r_1^2)$
	AaBB	R	$0.5r_1(1-r_1)$	$2r_1(1-r_1^2)/(2+r_1^2)$
	AaBb	R	$0.5(r_1)^2$	$2(r_1)^2/(2+r_1^2)$
	AabB	P	$0.5(1-r_1)^2$	$2(1-r_1)^2/(2+r_1^2)$
		Sum	$P_k=0.25(2+r_1^2)$	
(Ab)	AAbb	P	$0.25(1-r_1)^2$	$(1-r_1)^2/(1-r_1^2)$
	Aabb	R	$0.5r_1(1-r_1)$	$2r_1(1-r_1)/(1-r_1^2)$
		Sum	$P_k=0.25(1-r_1^2)$	
(aB)	aaBB	P	$0.25(1-r_1)^2$	$(1-r_1)^2/(1-r_1^2)$
	aaBb	R	$0.5r_1(1-r_1)$	$2r_1(1-r_1)/(1-r_1^2)$
		Sum	$P_k=0.25(1-r_1^2)$	
(ab)	aabb	R	$0.25r_1^2$	1

(r_1<0.05) because the surmised genotype for each phenotype except for the double recessive phenotype should be correct with P>0.90 (Table 1, Fig. 1).

In practice, the true value of r_1 is not known, and treating a TDLM pair as a "co-dominant megalocus" in an F_2 family depends upon the precision of the prior estimation of r_1. The prior estimate of r_1 for the F_1 individual is obtained from backcross-model segregation data for an additional family that also uses the F_1 parent. Even if no recombinant progeny are observed for a pair of markers in

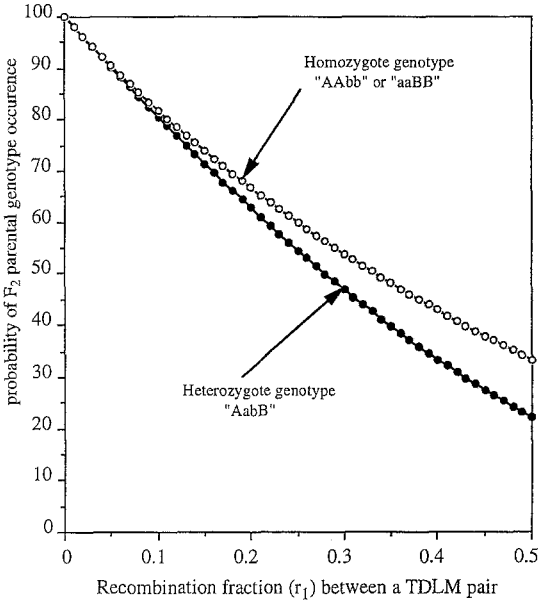


Fig. 1 Conditional probability of F_2 parental genotype occurrence for a TDLM pair, plotted against the recombination fraction r_1. The dominant marker phenotype of an F_2 progeny is assigned to the parental (non-recombinant) genotypic class. The probability of occurrence is computed from Table 1, as $P_{i,P}|P_k$ for the k^{th} phenotypic class, were $P_{i,P}|P_k$ is the conditional probability of F_2 parental genotypes, i, given the F_2 phenotype, k, for a TDLM pair

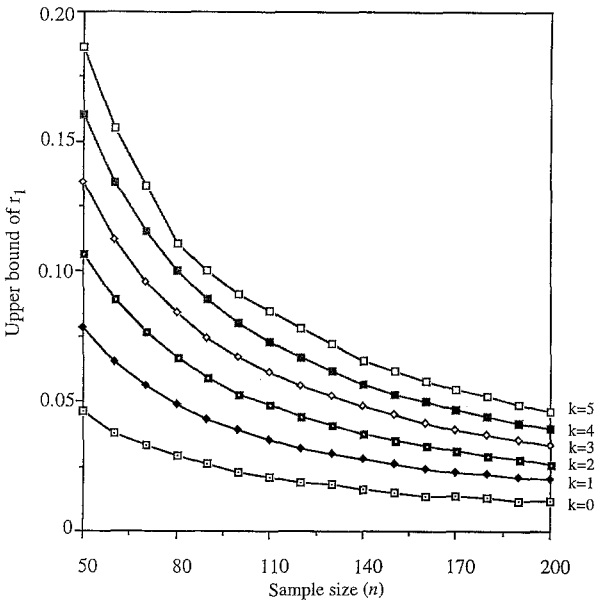
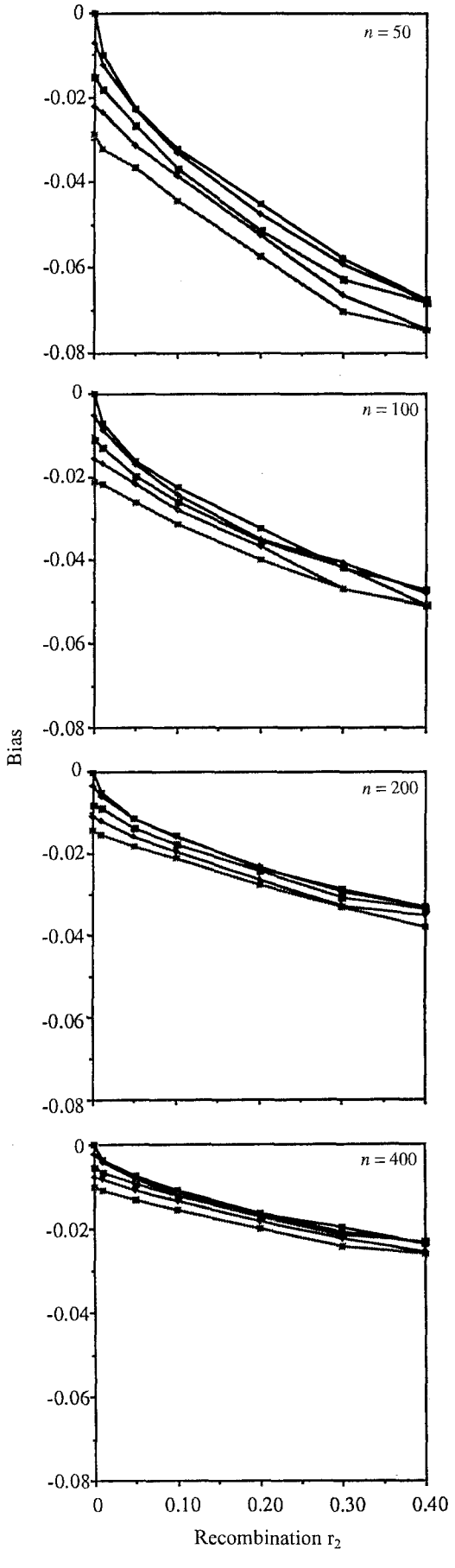


Fig. 2 Upper-bound estimate of the recombination fraction r_1 for the outcome of k recombinants (k=0 to 5) observed in a backcross-type population, plotted against the sample size n (n=50 to 200). The probability of observing as few as k recombinants is <0.10 for a recombination fraction greater than the upper bound

Table 2 Information function, $I(r_2|r_1)$, for a trans-dominant linked marker (TDLM) pair in an F_2 population, following Allard (1956). The recombination frequency for the interval A-B is r_1 and for the interval (AB)-C is r_2, $r=r_1+r_2$. A and B are two dominant markers while C is a co-dominant locus where C and c are dominant and recessive alleles. The expected frequencies (r_{2ij}) of (AB)-C genotypes are computed assuming no double crossovers. The conditional probabilities for recombination events given the joint genotype, $P_{ij}(R|G)$, are computed by assimilating the TDLM as a "co-dominant megalocus". $P(Q_j|\{AB\}_i)$ is the frequency of the j^{th} putative co-dominant QTL, C, conditional on the i^{th} TDLM genotype

TDLM phenotype	Third locus genotype	Expected frequency (r_{2ij})	$P_{ij}(R\|G)$	$\frac{1}{r_{2ij}}\left(\frac{\delta(r_{2ij})}{\delta r_2}\right)^2$	$P[Q_j\|(AB)_i]$
(AB)	CC	$0.5r_2(1-r_1-r_2)$	0.5	$(1-r_2-r)^2/2r_2(1-r)$	$r_2(1-r_2)$
	Cc	$0.5[1-r_1-2r_2+r_2^2+(r_1+r_2)^2]$	$r_2^2/[r_2^2+(1-r_2^2)]$	$2(1-r_2-r)^2/(1-r-r_2+r^2+r_2^2)$	$1-2r_2+2r_2^2$
	cc	$0.25[2r_1+2r_2-r_2^2-(r_1+r_2)^2]$	0.5	$(1-r_2-r)^2/(2r-r^2-r_2^2)$	$r_2(1-r_2)$
(Ab)	CC	$0.25(1-2r_2-r_1^2+r_2^2)$	0	$(r_2-1)^2/(1-r_1^2-2r_2+r_2^2)$	$(1-r_2)^2$
	Cc	$0.5r_2(1-r_2)$	0.5	$(1-2r_2)^2/2r_2(1-r_2)$	$2r_2(1-r_2)$
	cc	$0.25r_2^2$	1	1	r_2^2
(aB)	CC	$0.25r_2(2r_1+r_2)$	1	$r^2/(r^2-r_1^2)$	r_2^2
	Cc	$0.5(1-r_1-r_2)(r_1+r_2)$	0.5	$(1-2r)^2/2r(1-r)$	$2r_2(1-r_2)$
	cc	$0.25(1-r_1-r_2)^2$	0	1	$(1-r_2)^2$
(ab)	CC	$0.25r_1^2$	NA	0	NA
	Cc	NA	NA	NA	NA
	cc	NA	NA	NA	NA

$$I(r_2|r_1)=\sum_{ij}\frac{1}{r_{2ij}}\left(\frac{\delta(r_{2ij})}{\delta r_2}\right)^2$$



the backcross family, the true value of the recombination fraction r_1 could be greater than 0. To evaluate the effect of sample size on the estimation of r_1, we computed the upper-bound estimate of the recombination fraction for the outcome of k=0 to 5 recombinants observed in a backcross progeny (Fig. 2). Using the Poisson distribution, we numerically determined the value of r_1 such that the probability of observing as few as k recombinants was <0.10 to provide an approximate 90% upper bound for the estimate of r_1 (Fig. 2). Our results show that little precision is gained by increasing the sample size above 120. If no recombinants are observed for n>50, the probability of the true value of r_1 exceeding 4.5 cM is less than 0.10. If five or fewer recombinants are observed for n>120, the probability of the true value of r_1 exceeding 0.08 is less than 0.10.

Maximum-likelihood estimation of the recombination fraction r_2 between a TDLM pair and a co-dominant marker

To obtain a maximum-likelihood estimate (MLE) of the recombination fraction r_2, (denoted $\hat{r}_{2ML}$) the expected frequency, r_{2ij}, of each F_2 genotype was calculated for the "co-dominant megalocus" (AB) and the co-dominant locus C (Table 2). The MLE was computed from the conditional probabilites, $P_{ij}(R|G)$, for recombination events given each genotype using the expectation maximization (EM) algorithm (Dempster et al. 1977). We used simulation to estimate the bias of $\hat{r}_{2ML}$. The EM algorithm was applied to several numerical examples using expected values of the numbers of offspring in each phenotypic class given a sample size of n=50, 100, 200 and 400 individuals and several combinations of values for r_1 (0.00, 0.01, 0.05, 0.10 and 0.20) and r_2.(0.00, 0.01, 0.05, 0.10, 0.20, 0.30 and 0.40). A total of 1000 samples were simulated for each variable combination. Mean bias was estimated by $(r_2-\hat{r}_{2ML})$, with r_2 the true recombination frequency and $\hat{r}_{2ML}$ the mean MLE of r_2 calculated from the 1000 samples. For each combination, the bias was different from zero and consistently more than r_2. Thus $\hat{r}_{2ML}$ was an upwardly biased estimator of r_2. This bias decreased as r_1 and r_2 decreased and the sample size increased (Fig. 3). However, the bias was not greatly affected by the different values of r_1.

The variance of the MLE of r_2 (given r_1) in an F_2 family was evaluated as an information function, $I(r_2|r_1)$. This information function had a complex expression (Table 2). We plotted $I(r_2|r_1)$ relative to the backcross information function for $0 \le r_1 \le 0.15$ (Fig. 4). When r_1=0, $I(r_2|r_1)$ was equal to the information function for a pair of co-dominant markers. For r_1=0.01, our result showed that the information function for the MLE of the recombination fraction,

Fig. 3 Bias of $\hat{r}_{2ML}$ estimated from a simulation for recombination frequencies between a TDLM pair and a co-dominant locus, r_2, ranging from 0.00 to 0.40, with sample sizes of 50, 100, 200 and a 400 F_2 progeny and a recombination frequency within a TDLM pair of r_1 . The different curves correspond to different r_1 taking values of of 0.00 (upper curves), 0.01, 0.05, 0.10 and 0.20 (lower curve)

r_2, between a TDLM pair and a co-dominant marker was almost identical with two co-dominant markers when $r_2>0.05$. In all cases this model provided much more information than dominant markers alone in either a cis or trans configuration (Fig. 4). For r_1 ranging between 0.05 and 0.15, a TDLM pair provided 75%–66% of the information content of two co-dominant markers (Fig. 4). For large values of r_2, the TDLM model information rose above that for F_2 co-dominant markers, but the assumptions used were not realistic in this range of recombination values.

Estimation of additive and dominance effects for a QTL in F_2 families using dominant and co-dominant markers

A megalocus approximation of co-dominant markers in an F_2 family can be justified by an examination of the analysis of variance for estimating gene dosage effects. Marker genotypes serve as classification variables to test for the association of trait values with genetic markers. In the case of a dominant marker and a co-dominant QTL, only one contrast can be tested which is: $\mu_M-\mu_{mm}=0$, where, μ_M is the mean value for dominant marker phenotypes M (genotype MM or Mm) and μ_{mm} is the mean value for marker genotypes mm. The expectation of this contrast is: $\frac{3}{4}(1-2r_2)a+\frac{2}{3}(1-2r_2)^2d$, where a and d are the additive and dominance genetic effects of the QTL, respectively. Thus, this contrast tests for confounded additive and dominance effects and the linkage between the marker and the putative QTL. When co-dominant markers are used, both additive and dominance effects can be tested separately (Edwards et al. 1987). For this reason, Paterson et al. (1991b) identified F_2 self families as the "ideal" population for studying gene action. When TDLM pairs are used, the expectations for the two contrasts used to evaluate gene action can be derived from Table 3 and are expressed as:

$$E(\text{"additive"}) = \frac{\mu_{LL}-\mu_{ll}}{2} = (1-2r_2)a \approx \frac{\mu_{Ab}-\mu_{aB}}{2}$$

$$E(\text{"dominant"}) = \mu_{Ll} - \frac{\mu_{LL}+\mu_{ll}}{2}$$

$$= (1-2r_2)^2 d \approx \mu_{AB} - \frac{\mu_{Ab}+\mu_{aB}}{2}.$$

If A and B are tightly linked, the hypothesis tests on the contrasts have a biological meaning for the putative QTL. Thus, even though a "co-dominant megalocus" introduces some loss of precision compared with real co-dominant markers, there is a substantial advantage to re-coding TDLM pairs as co-dominant loci for detecting QTLs and studying gene action in an F_2 family.

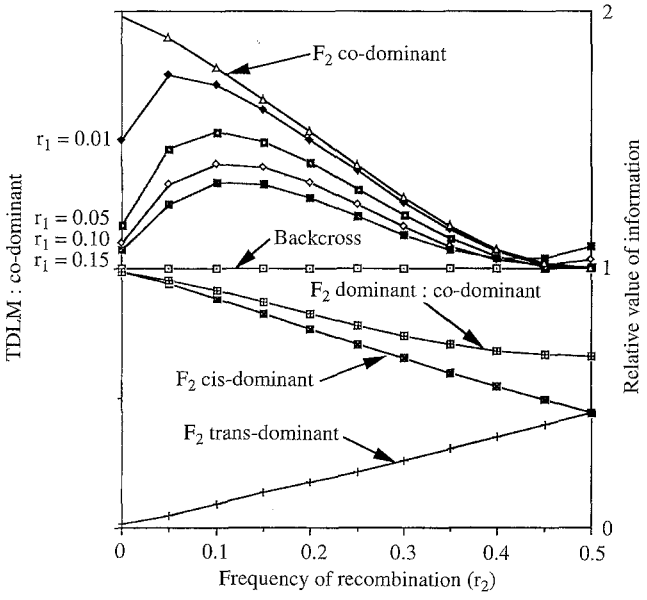


Fig. 4 Information functions for a TDLM pair linked to a co-dominant marker, compared with four other marker-configurations relative to the information function of the backcross population. Information for the TDLM model are computed for several values of r_1 (r_1=0.01, 0.05, 0.1, 0.15)

Discussion

Dominant markers are easily mapped in backcross families, in haploid families (doubled-haploid or conifer megagametophytes), or in recombinant inbred lines. F_2 families offer significant advantages for the genetic analysis of QTLs, but dominant markers provide no information on gene dosage effects and dominant markers are seldom used for the genetic analysis in F_2 families. Currently, no mapping software offers algorithms for quantitative trait dissection in F_2 progenies using dominant markers. Here we have considered how prior knowledge of linkage arrangements in the F_1 parent obtained from backcross-type progeny could be used to assemble pairs of dominant markers

Table 3 TDLM genotypes, their corresponding co-dominant marker genotypes, their marginal frequencies and their expected trait values. a and d are the additive and dominance effects as defined by Edwards et al. (1987), respectively. L and l are dominant and recessive alleles at a re-coded co-dominant locus, respectively

TDLM phenotype	TDLM expected frequency	TDLM recoded as a co-dominant locus	TDLM approximate frequency when $r_1 \to 0$	Expected trait value of a co-dominant marker (Edwards et al. 1987)
(AB)	$0.25(2+r_1^2)$	Ll	0.5	$\mu_{Ll} = [(1-r_2)^2 + r_2^2]d$
(Ab)	$0.25(1-r_1^2)$	LL	0.25	$\mu_{LL} = (1-2r_2)a + 2r_2(1-r_2)d$
(aB)	$0.25(1-r_1^2)$	ll	0.25	$\mu_{ll} = -(1-2r_2)a + 2r_2(1-r_2)d$
(ab)	$0.25r_1^2$	NA		

into co-dominant "megaloci" for the genetic analysis of traits in an F_2 family.

Pairs of dominant markers that are tightly linked in repulsion (TDLMs) provide an approach to genotypic inference for F_2 progenies based upon conditional probabilities. If the recombination fraction between the two markers is small, the conditional probability of observing non-recombinant classes, given the F_2 phenotype, is very large. When the recombination fraction approaches zero, the dominant marker genotype of an F_2 progeny can be surmised with near certainty from the progeny TDLM phenotype. The accuracy of the surmised F_2 genotypes depends upon the precision of the estimate for the recombination fraction for the TDLM pair. We showed that the upper bound of the estimate of the recombination fraction for a TDLM decreased with increasing sample size, but that little precision was gained by raising the "backcross" sample size above 120. In an F_2 family, the three genotypic classes of a single co-dominant marker correspond closely with the three most probable TDLM genotypic classes. The last TDLM genotypic class (aabb) was the only one where recombinant gametes can be inferred without ambiguity. However, this class has a very small chance of being observed (Knapp et al. 1995). Maximum-likelihood estimation of the recombination fraction between a TDLM pair and a co-dominant locus in an F_2 family provides an estimator with small bias and with linkage information almost equivalent to two co-dominant loci when the map distance within the TDLM pair is small. Compared to dominant markers, co-dominant markers have the advantage for genetic analysis that the tests for QTL genetic effects (additive and dominant) are not confounded. Treating dominant markers tightly linked in repulsion provides a substantial benefit for genetic analysis in an F_2.

The value of the TDLM approach depends upon the relative advantages of different systems of molecular markers. These relative advantages differ for different plants. For example, co-dominant RFLP methods could be favored if mapped probes were readily available (e.g. *Arabidopsis*, Chang et al. 1988; Nam et al. 1989). On the other hand, a saturated genetic map of RAPD markers was constructed with great efficiency and speed in an *Arabidopsis* recombinant inbred (RI) population (Reiter et al. 1992). Several generations of inbreeding are required to develop RI populations, thus the RI approach costs time and is not feasible for some species. In contrast to the RI approach, the TDLM approach allows dominant markers to be used immediately for genetic analysis in F_2 families. These methods have special advantages for allogamous species such as forest trees where little prior genetic information is available for many species and generation times are long. The TDLM model is an alternative to laboratory approaches for converting dominant markers into co-dominant markers. Distinguishing between RAPD bands from heterozygous and homozygous genotypes could be based on intensity differences, but quantitation is difficult and the general utility of this approach remains to be determined (Williams et al. 1993). RAPD marker fragments have been cloned and sequenced, allowing oligonucleotide primers to be designed that specifically PCR-amplify the corresponding genomic DNA sequence (Paran and Michelmore 1993). These sequence-tagged sites for RAPD markers called SCARs (sequenced characterized amplified regions), can be scored as co-dominant loci. SCARs can be digested with restriction enzymes that can reveal internal nucleotide sequence polymorphisms. Converting dominant markers to SCARs for within-family genetic analysis could also be justified if the cost to obtain as many as two-fold more data points for TDLM sites is greater than the cost for cloning and sequencing many dominant markers. SCARs could also be useful in some other crosses where one or both parents are heterozygous.

Genotypes for large numbers of progeny are necessary for QTL detection, but only a small number of co-dominant markers are needed from each linkage group (Darvasi et al. 1993). If the number of PCR reactions for QTL analysis is limiting, then more individuals can be assayed if fewer sites per linkage group are chosen. Only half as many PCR reactions would be required to resolve co-dominant microsatellite markers compared to RAPD markers. Microsatellites could mark homologous locations for different mapping populations, thus providing mapping information that is readily transferable among individuals. However, microsatellites have significant development costs and they must be resolved on DNA sequencing gels, a more laborious method compared with RAPDs. AFLPs are also resolved on sequencing gels, but many more data points are obtained per PCR reaction and per sequencing gel, compared with microsatellites. TDLM sites should be abundant on the genetic map of F_1 individuals, so F_2 families could be analyzed using dominant AFLPs. Note that a few TDLMs evenly spaced in each linkage group will be needed to scan the genome for marker-trait association. While the full set of AFLP mapping reactions and gels could be needed for each sample to obtain those markers linked in trans, RAPDs could be more selective than AFLPs because only those primers that give informative markers would be selected. Saturated maps have been constructed for some plant species using RFLPs (Tanksley et al. 1992; Vallejos et al. 1992), but RAPDs and AFLPs offer significant advantages in speed and efficiency for constructing saturated maps to locate important genes, especially in breeding populations (Al-Janabi et al. 1993; Grattapaglia and Sederoff 1994; Becker et al. 1995; Plomion et al. 1995; van Eck et al. 1995).

We used the maritime pine genetic system and a previously reported genetic map based on 436 RAPD markers assayed on 124 megagametophytes (Plomion et al. 1995) to evaluate the distribution of potential TDLM sites. Fifty seven TDLM pairs ($r_1<0.05$ cM) covering 72.5% of the genome were found. Each megagametophyte (a haploid tissue in seeds that is derived from the same megaspore that gives rise to the maternal gametes) corresponding to each F_2 plant was collected a few weeks after germination and just before the seed coat would be cast off. The plants were grown and mesured for quantitative traits. The purpose of this "backcross map" was to identify RAPD markers tightly linked in repulsion to the genotype of the F_2 seedlings for

detecting QTLs and studying gene action with dominant markers. This map information shows that the TDLM approach applied to RAPD markers is a useful way to obtain genome coverage for an analysis of traits in an F_2 family.

Acknowledgements This work was partly supported by grants from France (AIP Génome des Plantes Cultivées), from the U.S. Department of Agriculture (NRICGP No. 92-37300-7549, 93-37300-8839) and the USDA Forest Service (Coop Agreement No. 29-1009). We wish to thank Ron Sederoff for inviting C.P. to conduct research with the Forest Biotechnology Group, North Carolina State University, Raleigh.

References

Al-Janabi SM, Honeycutt RJ, McClelland M, Sobral BWS (1993) A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. Genetics 134:1249–1260

Allard R (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235–278

Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249:65–73

Caetano-Anolles G, Bassam BJ, Gresshoff PM (1991) A high-resolution DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Biotechnology 9:553–557

Chang C, Bowman AW, Lander ES, Meyerowitz EW (1988) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85:9856–6860

Chaparro JX, Werner DJ, O'Malley D, Sederoff RR (1994) Targeted mapping and linkage analysis of morphological, isozyme and RAPD markers in peach. Theor Appl Genet 87:805–815

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

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Statist Soc Series B 39:1–39

Eck HJ van, van der Voort JR, Draaistra J, van Zandvoort P, van Enckevort E, Segers B, Peleman J, Jacobsen E, Helder J, Bakkler J (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. Mol Breed 1:397–410

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilited investigations of quantitative trait loci in maize. I Numbers, genomic distribution and type of gene action. Genetics 116:113–125

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

Knapp SJ, Holloway JL, Bridges WC, Liu BH (1995) Mapping dominant markers using F_2 matings. Theor Appl Genet 91:74–81

Nam HG, Giraudat J, den Boer B, Moonan F, Loos WDB, Hauge BM, Goodman HM (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. Plant Cell 1:699–705

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85:985–993

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991a) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics 127:181–197

Paterson AH, Tanksley SD, Sorrells ME (1991b) DNA markers in plant improvement. Adv Agron 46:39–90

Plomion C, Bahrman N, Durel C-E, O'Malley DM (1995) Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. Heredity 74:661–668

Reiter RS, Williams J, Feldman K, Rafalski JA, Tingey SV, Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89:1477–1481

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131:733–740

Vos P, Hogers R, Bleeker M, Reijans M, van der Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. Methods Enzymol 218:704–740